

UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE CIÊNCIAS EXATAS E DA TERRA
PROGRAMA DE PÓS-GRADUAÇÃO EM DEMOGRAFIA

WALTER PEDRO SILVA JÚNIOR

**DIFERENCIAIS REGIONAIS NA MORTALIDADE ADULTA POR
ESCOLARIDADE NO BRASIL EM 2010**

NATAL/RN

2018

WALTER PEDRO SILVA JÚNIOR

**DIFERENCIAIS REGIONAIS NA MORTALIDADE ADULTA POR
ESCOLARIDADE NO BRASIL EM 2010**

Dissertação apresentada ao Programa de Pós-graduação em Demografia da Universidade Federal do Rio Grande do Norte (UFRN), como requisito parcial para obtenção do título de mestre em Demografia.

Orientador: Prof. Dr. Flávio Henrique M. de A. Freire
Coorientador: Prof. Dr. Damião Nóbrega da Silva

NATAL/RN

2018

FICHA CATALOGRÁFICA

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Central Zila Mamede

Silva Junior, Walter Pedro.

Diferenciais regionais na mortalidade adulta por escolaridade no Brasil em 2010 / Walter Pedro Silva Junior. - 2018.
110f.: il.

Dissertação (Mestrado)-Universidade Federal do Rio Grande do Norte, Centro de Ciências Exatas e da Terra, Programa de Pós-graduação em Demografia, Natal, 2018.

Orientador: Dr. Flávio Henrique Miranda de Araújo Freire.

Coorientador: Dr. Damião Nóbrega da Silva.

1. Mortalidade - Dissertação. 2. Nível de escolaridade - Dissertação. 3. Diferencial Regional - Dissertação. 4. Dados faltantes - Dissertação. I. Freire, Flávio Henrique Miranda de Araújo. II. Silva, Damião Nóbrega da. III. Título.

RN/UF/BCZM

CDU 314

Elaborado por Raimundo Muniz de Oliveira - CRB-15/429

WALTER PEDRO SILVA JÚNIOR

**DIFERENCIAIS REGIONAIS NA MORTALIDADE ADULTA POR
ESCOLARIDADE NO BRASIL EM 2010**

Prof. Dr. Flávio Henrique Miranda de Araújo Freire
Universidade Federal do Rio Grande do Norte - UFRN
(Orientador)

Prof. Dr. Damião Nóbrega da Silva
Universidade Federal do Rio Grande do Norte - UFRN
(Coorientador)

Prof. Dr. Marcos Roberto Gonzaga
Universidade Federal do Rio Grande do Norte - UFRN
(Membro Interno)

Prof. Dr. Cássio Maldonado Turra
Centro de Desenvolvimento e Planejamento Regional - CEDEPLAR/UFMG
(Membro Externo)

Resultado _____

*Dedico aos meus pais Isailda Alves
e Walter Pedro que, além de me
proporcionarem amor e educação,
ensinaram-me o caminho do bem.*

AGRADECIMENTOS

A Deus pelo dom da vida e por ter me capacitado nos momentos mais difíceis desta conquista.

Ao meu filho, Pedro Caio, por ser esse anjo em minha vida. Em meio às turbulências desta jornada, eis que Deus me presenteia com tua chegada. A partir de então, és minha maior motivação para conclusão dos meus projetos de vida, meu amor.

Aos meus amados pais, Walter Pedro e Isailda Alves, e ao meu querido irmão e melhor amigo, Walker Alves, pelo amor, apoio e confiança em todos os momentos. Peço perdão pelos momentos em que me ausentei de suas vidas em busca desse objetivo.

À Maíra Costa dos Santos pela paciência no convívio diário nesses quase dois anos. Muito obrigado pelo companheirismo e amor, mesmo diante das dificuldades. Você tem uma parcela considerável nesta conquista.

Ao meu orientador professor Flávio Henrique Miranda de Araújo Freire por ser o grande responsável pela realização deste trabalho. Tive a sorte de ter o senhor como orientador. Muito obrigado pelo acolhimento e dedicação nas reuniões de orientação. A cada conversa eu saía motivado a dar continuidade ao trabalho na certeza de que estava no rumo certo. Isso me ajudou demais. Muito obrigado por tudo e parabéns pelo profissional e pessoa que és!

Ao meu coorientador professor Damião Nóbrega da Silva pelos ensinamentos e contribuições sobre os métodos de imputação de dados utilizados neste trabalho. Foi um desafio muito grande para mim, era tudo novo, mas ficou mais tranquilo com seus direcionamentos.

Aos professores Marcos Roberto Gonzaga e Cássio Maldonado Turra pela imensa contribuição na elaboração final desta dissertação. Foi muito gratificante ter duas referências em estudos sobre mortalidade na minha banca de defesa. Minha gratidão aos senhores.

Aos colegas da coorte de 2017. Vocês tornaram essa “digressão” mais agradável e divertida. Em especial, a João, Rafael e Albéria pela parceria. Sou grato pela amizade de vocês.

A todos os professores, técnicos, bolsistas e terceirizados do PPGDEM e do DDCA da UFRN. Vocês contribuíram muito para meu crescimento profissional. Em especial ao aluno e bolsista Danilo pela contribuição no desafio que foi “rodar” todo meu trabalho no software R.

À Universidade Federal do Rio Grande do Norte, em especial à Pró-reitoria de Gestão de Pessoas (nas pessoas de Mirian Dantas dos Santos e Joade Cortez Gomes), pela oportunidade

de poder me capacitar e por contribuírem com meu crescimento profissional. Espero retribuir todo o investimento com um serviço de qualidade prestado à instituição e à sociedade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) pelo apoio concedido aos Programas de Pós-graduação do Brasil, em especial ao PPGDEM.

“Talvez não tenha conseguido fazer o melhor, mas lutei para que o melhor fosse feito. Não sou o que deveria ser, mas Graças a Deus, não sou o que era antes”.
(Marthin Luther King)

RESUMO

Estudos sobre diferenciais de mortalidade por escolaridade são de grande relevância para a gestão da saúde pública no Brasil. Dado o contexto desigual dos processos de transição demográfica e epidemiológica entre estratos sociais, o objetivo desse trabalho é estimar os diferenciais educacionais na mortalidade adulta (25 a 59 anos) por grandes regiões, sexo e idade em 2010. A literatura tem apontado para uma relação inversa entre a mortalidade e a escolaridade, ou seja, quanto maior a escolaridade menor é o risco de o indivíduo morrer. Além disso, há evidências de que esses gradientes persistem em todas as idades, são maiores entre os homens, diminuem com a idade e sua magnitude difere entre as grandes regiões do país. Enquanto essa temática vem sendo discutida nos Estados Unidos desde a década de 1960, no Brasil os primeiros trabalhos são da década de 2000. Esse atraso é devido, principalmente, à péssima qualidade da variável escolaridade das declarações de óbitos registradas no Sistema de Informação sobre Mortalidade (SIM). Para buscar evidências sobre esses gradientes educacionais, os microdados de óbitos do SIM em 2010 (para os óbitos) e os microdados do censo 2010 (para as quantidades populacionais) são as fontes de dados do estudo. Com isso, a proposta metodológica neste trabalho é realizar duas correções nos dados antes de proceder com as estimativas dos diferenciais educacionais na mortalidade. A primeira é corrigir a incompletitude da variável escolaridade através de métodos de imputação. Os métodos aplicados são: imputação múltipla e ABB – *Approximate Bayesian Bootstrap*. A segunda trata-se da correção de sub-registro dos óbitos, utilizando fatores de correção médios da população geral por UF para corrigir apenas os óbitos de menores escolaridades em cada região. Os resultados em termos das taxas específicas de mortalidade estão em consonância com a literatura. Os riscos de morte são menores entre os indivíduos de alta escolaridade comparados aos de baixa e média escolaridade. Além disso, os diferenciais foram maiores para os homens e tendem a diminuir com a idade. Outro resultado importante é que não há diferenças regionais nos diferenciais de mortalidade segundo a escolaridade do falecido.

Palavras-chave: Mortalidade. Nível de escolaridade. Diferencial Regional. Dados faltantes

ABSTRACT

Studies on mortality differentials by education have a great relevance for the management of public health in Brazil. Given the unequal context of processes of demographic and epidemiological transitions among social strata, the objective of this study is to estimate educational differentials in adult mortality (25-59 years) by regions, sex and age in 2010. The literature indicates that there is an inverse relationship between mortality and schooling, meaning that the higher the schooling the lower is the risk of death. In addition, there is evidence that these gradients persist at all ages. They are larger among men, diminishing with age and their magnitude differ among the country's regions. Whilst this subject has been discussed in the United States since the 1960s, in Brazil the first works date from the 2000s. This delay is mainly due to the poor quality of the educational variable from the responses on death registered in the Mortality Information System (MIS). In order to search for evidence on these educational gradients, the microdata on death of MIS of 2010 (for deaths) and the microdata of the census 2010 (for population quantities) are the study's sources of data. Accordingly, the methodological proposal is to make two corrections in the data before proceeding with the estimates of educational differentials in mortality. The first one is the correction of the incompleteness of the variable education through imputation methods. The methods applied are: the multiple imputation and the ABB (Approximate Bayesian Bootstrap). The second one is the correction of underreporting of deaths using average corrections factors of the general population by Federation Unit (FU) to correct only the deaths of lower schooling in each region. The results in terms of specific mortality rates are in line with the literature. The risks of death are lower among individuals with high schooling when compared to those with low and medium schooling. Moreover, differentials were higher for men, tending to decrease with age. Another important result is that there are no regional differences in the mortality differentials according to the schooling of the deceased.

Keywords: Mortality. Level of schooling. Regional differential. Missing data.

LISTA DE FIGURAS

Figura 1: Participação (%) da população residente nas UFs da região Norte por situação de domicílio rural, 2004 a 2015.	43
Figura 2: Fluxo operacional para obtenção e consolidação da base de dados de óbitos registrados pelo Ministério da Saúde/ Datasus / SIM em 2010.	44
Figura 3: Percentuais de dados faltantes na variável escolaridade do óbito de 25 a 59 anos incluindo os casos com a idade faltante, Brasil, grandes regiões, 1979 a 2016.	50
Figura 4: Tipos de padrão de dados faltantes.	61
Figura 5: Ilustração do método ABB com um caso hipotético.	64
Figura 6: Padrões dos dados faltantes nos registros de óbitos do SIM no Brasil em 2010, segundo variáveis sociodemográficas.	70
Figura 7: Distribuição dos óbitos segundo os 10 padrões mais frequentes no conjunto de dados.	71
Figura 8: Logaritmo das taxas específicas de mortalidade por nível de escolaridade segundo região e sexo no cenário com correção de sub-registro, 2010.	80
Figura 9: Logaritmo das taxas específicas de mortalidade por nível de escolaridade segundo região e sexo no cenário sem correção de sub-registro, 2010.	81
Figura 10: Razão de risco por idade segundo o sexo e grandes regiões no cenário com correção de sub-registro, 2010.	83
Figura 11: Razão de risco por idade segundo o sexo e grandes regiões no cenário sem correção de sub-registro, 2010.	83
Figura 12: Razão de risco (RR) segundo o sexo por idade e grandes regiões no cenário com correção de sub-registro, 2010.	84
Figura 13: Razão de risco (RR) segundo o sexo por idade e grandes regiões no cenário sem correção de sub-registro, 2010.	85
Figura 14: Razão de risco (RR) por regiões, segundo sexo e idade com correção de sub-registro, 2010.	86
Figura 15: Razão de risco (RR) por regiões, segundo sexo e idade sem correção de sub-registro, 2010.	87

LISTA DE TABELAS

Tabela 1: Variáveis sociodemográficas dos óbitos registrados no SIM em 2010 e categorizações utilizadas.....	44
Tabela 2: Distribuição do número de óbitos adultos no Brasil em 2010 segundo variáveis demográficas e escolaridade.....	51
Tabela 3: Distribuição e Razões de Chance (RC) da variável Escolaridade ignorada dos óbitos adultos segundo variáveis sociodemográficas, Brasil, 2010.	53
Tabela 4: Frequência absoluta e relativa de estratos segundo a condição da quantidade de doadores disponíveis para imputar escolaridade do óbito.....	69
Tabela 5: Comparação entre a distribuição de óbitos por idade com e sem imputação segundo sexo e escolaridade, Brasil, 2010.	72
Tabela 6: Comparação entre a distribuição de óbitos com e sem imputação segundo sexo e escolaridade, Brasil, 2010.	72
Tabela 7: Variáveis sociodemográficas do censo demográfico 2010 utilizadas.	75
Tabela 8: Compatibilização entre a variável “Anos de estudo” do SIM 2010 e “nível de instrução” do Censo 2010.	76

LISTA DE ABREVIATURAS E SIGLAS

AASCR	Amostra Aleatória Simples Com Reposição
ABB	<i>Approximate Bayesian Bootstrap</i>
BIC	<i>Bayesian Information Criterion</i>
CID	Classificação Internacional de Doenças
CPS	<i>Current Population Survey</i>
DO	Declaração de Óbito
EM	<i>Expectation-Maximization</i>
ER	Eficiência Relativa
GGB	<i>General Growth Balance</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IM	Imputação Múltipla
IRD	Índice Relativo de Desigualdade
MAR	<i>Missing At Random</i>
MCAR	<i>Missing Completely At Random</i>
MCD	<i>Multiple Cause of Death</i>
MCMC	Monte Carlo via Cadeia de Markov
MS	Ministério da Saúde
MV	Máxima Verossimilhança
MVN	Modelo Multivariado Normal
NCHS	<i>National Center for Health Statistics</i>
NDI	<i>National Death Index</i>
NHIS	<i>National Health Interview Survey</i>
NLMS	<i>National Longitudinal Mortality Study</i>
NMAR	<i>Not Missing At Random</i>
PIA	População em Idade Ativa
PNAD	Pesquisa Nacional por Amostra de Domicílios
PPV	Pesquisa sobre Padrões de Vida
RC	Razão de Chance

RR	Razão de Risco
RP	Razão de Probabilidade
SEG	Gerações Extintas Sintético
SEG Adj	Gerações Extintas Sintético Ajustado
SIM	Sistema de Informação sobre Mortalidade
SSE	<i>Status</i> Socioeconômico
SUS	Sistema Único de Saúde
TEM	Taxa Específica de Mortalidade
UF	Unidade da Federação

SUMÁRIO

1 INTRODUÇÃO	16
1.1 Objetivos	24
1.1.1 Objetivo geral	24
1.1.2 Objetivos específicos.....	24
2 DIFERENCIAIS DE MORTALIDADE: REVISÃO DE LITERATURA	25
3 AVALIAÇÃO DA QUALIDADE DOS DADOS: COMPLETITUDE	39
3.1 Metodologia	41
3.1.1 Fonte de dados.....	41
3.1.2 Método	48
3.2 Resultados	49
3.3 Discussão.....	54
4 IMPUTAÇÃO DE DADOS FALTANTES NA ESCOLARIDADE DO ÓBITO NO SIM. 57	
4.1 Metodologia	59
4.1.1 Fonte de dados.....	59
4.1.2 Método	60
4.1.2.1 Padrões e mecanismos geradores de dados faltantes	60
4.1.2.2 Métodos de imputação	62
4.1.2.3 Seleção das variáveis para o método Hot Deck com seleção via <i>ABB</i>	67
4.2 Resultados da imputação	69
4.3 Discussão.....	73
5 DIFERENCIAIS DE MORTALIDADE ADULTA NO BRASIL, SEGUNDO ESCOLARIDADE	75
5.1 Metodologia	75
5.1.1 Fonte de dados.....	75
5.1.2 Ajustes nos dados	77
5.2 Resultados	79
6 CONSIDERAÇÕES FINAIS	89
REFERÊNCIAS	92
APÊNDICE	103

1 INTRODUÇÃO

Sabe-se que compreender o comportamento dos diferenciais¹ de mortalidade adulta segundo o *status* socioeconômico via dimensão escolaridade é de suma importância para o monitoramento e elaboração de políticas mais eficientes, eficazes e inclusivas no âmbito da saúde pública e de reformas no sistema previdenciário. Assim, este trabalho tem como objetivo estimar os diferenciais da mortalidade adulta no Brasil e regiões a partir da imputação dos dados de escolaridade do óbito registrados em 2010 pelo Sistema de Informação sobre Mortalidade (SIM).

Essa necessidade de se investigar essas desigualdades socioeconômicas na saúde via mortalidade é ainda mais essencial dado as experiências desiguais que as sociedades vivenciaram e vivenciam em termos da transição demográfica (NOTESTEIN, 1945; COALE, 1979; COALE, 1989; WOOD; CARVALHO, 1994; BRITO, 2008; VASCONCELOS; GOMES, 2012) e epidemiológica (OMRAN, 1971; LERNER, 1973; FRENK et al., 1991a; FRENK et al., 1991b; PRATA, 1992; CALDWELL, 1993; OMRAN, 1998; SCHRAMM et al., 2004; BORGES, 2017).

A transição demográfica é o processo no qual uma população passa de altas taxas de mortalidade e fecundidade para baixo níveis dessas componentes. O primeiro movimento ocorre com a queda das taxas de mortalidade, principalmente a mortalidade infantil. Como consequência, altas taxas de crescimento populacional foram observadas durante essa fase. No segundo momento, a fecundidade também declina diminuindo o crescimento populacional e alterando a estrutura etária (NOTESTEIN, 1945; COALE, 1979).

É importante ressaltar que esse processo ocorreu em ritmos e períodos diferentes, dadas as especificidades das populações. Por exemplo, na maioria dos países da Europa (países ditos desenvolvidos), a transição iniciou em meados do século XVIII e durou mais de um século (COALE, 1989). Por outro lado, no Brasil (país em desenvolvimento) esse processo se inicia em meados da década de 1940 com a queda da mortalidade. A partir da década de 1970, os declínios nos níveis de mortalidade e, principalmente da fecundidade se intensificaram. Apesar de ainda estarem em queda, aproximadamente 40 anos depois, essas componentes demográficas atingiram níveis consideravelmente mais baixos. Em 2010, a expectativa de vida brasileira ultrapassou os 70 anos e a fecundidade estava abaixo do nível de reposição (2,1 filhos por

¹ Diferenciais ou gradientes referem-se às diferenças observadas em uma medida de mortalidade entre as categorias de variáveis populacionais.

mulher em média), sinalizando uma transição bastante avançada (VASCONCELOS; GOMES, 2012).

Uma das principais consequências da transição demográfica é a mudança da estrutura etária. No Brasil, quando ocorre o declínio da mortalidade infantil, a participação da população jovem de 0 a 14 anos e a razão de dependência total aumentam, ficando em torno de 40% e 80%, respectivamente, até a década de 1970. A partir desse período, impulsionado pelo declínio da fecundidade, observa-se um decaimento da participação da população jovem e da razão de dependência. Ao mesmo tempo, as proporções da PIA - População em Idade Ativa (15 a 64 anos) e da população Idosa (65 ou mais) aumentaram, o que caracteriza um processo de envelhecimento da população brasileira (PAIVA; WAJNMAN, 2005; CARVALHO; RODRÍGUEZ-WONG, 2008). Vasconcelos e Gomes (2012) mostram que as regiões Norte, Nordeste e Centro-Oeste estavam mais atrasadas em relação ao Sudeste e Sul nesse processo de envelhecimento em 2010. Os índices de envelhecimento para as regiões Norte, Nordeste e Centro-Oeste foram de 24,6%, 38,6% e 36,0%, respectivamente. Já para as regiões Sudeste e Sul os índices foram 54,6% e 55,0%.

Baseado em projeções realizadas pelo IBGE, Carvalho e Rodríguez-Wong (2008) indicaram uma razão de dependência em seu menor patamar por volta do ano de 2025, o que significa que a PIA terá seu maior volume. Esse menor grau de dependência por parte dos jovens e idosos da PIA é benéfico no sentido de impulsionar o desenvolvimento econômico e melhorias nas condições de vida da população (ALVES, 2008). Esse fenômeno é denominado de Bônus Demográfico ou Janela de oportunidade (PAIVA; WAJNMAN, 2005; ALVES, 2008; BRITO, 2008; CARVALHO; RODRÍGUEZ-WONG, 2008; VASCONCELOS; GOMES, 2012). Como a dinâmica populacional brasileira é diferenciada, as regiões vivenciarão esse benefício em momentos distintos.

No que se refere à transição da mortalidade, essas desigualdades regionais e socioeconômicas diminuíram, mas continuam sendo observadas desde o início de seu declínio, mesmo que a mortalidade tenha alcançado baixos níveis de uma maneira geral no Brasil (WOOD; CARVALHO, 1994; BRITO, 2008; VASCONCELOS; GOMES, 2012). Nas primeiras décadas da transição já havia evidências de que os ganhos de anos de vida eram diferenciados entre as regiões brasileiras. Por exemplo, a expectativa de vida ao nascer na porção central da região Nordeste era 18 anos menor, quando comparada à região Sul no período de 1970-1980. (WOOD; CARVALHO, 1994). Mais recentemente, Vasconcelos e Gomes (2012) evidenciaram que a mortalidade ficou muito mais homogênea entre as regiões,

porém essas desigualdades ainda persistem mesmo que em menor grau. Em 2010, as maiores diferenças entre as expectativas de vida ao nascer observadas pelas autoras foram entre as regiões Nordeste e Norte em relação à região Sul, com diferenças de 4,71 e 3,04 anos, respectivamente.

Por um lado, as condições de saúde (medidas através da mortalidade) melhoraram ao longo do tempo e reduziram esses diferenciais devido a melhorias da cobertura dos sistemas de saneamento básico e ampliação do acesso a serviços de saúde pública como atenção à saúde e vacinação (VASCONCELOS; GOMES, 2012). Por outro, as diferenças persistem porque, historicamente, houve um acentuado desequilíbrio no desenvolvimento econômico, industrial, urbano e social entre as regiões (BRITO, 2008; VASCONCELOS; GOMES, 2012).

Todo esse cenário de mudanças demográficas traz consigo outro fenômeno importantíssimo para a contextualização de estudos sobre diferenciais de mortalidade por escolaridade. Trata-se da transição epidemiológica. Omran (1971) conceitua transição epidemiológica como um processo de mudanças nos padrões de saúde e doença das populações, levando em consideração a interação desses padrões com seus determinantes e suas consequências demográficas e socioeconômicas. De maneira geral, o autor define como padrão inicial uma predominância de doenças infecciosas, transitando para uma maior participação das doenças crônico-degenerativas e por doenças provocadas pelo homem ao final.

Esse processo é descrito em três etapas (OMRAN, 1971) e, posteriormente, outras duas são acrescentadas a essa abordagem (OMRAN, 1998). Inicia-se com a era da pestilência e da fome caracterizada por altas taxas de mortalidade e baixo crescimento populacional são observadas. A segunda etapa refere-se a era de recuo pandêmico, refletindo queda progressiva da mortalidade em função da menor ocorrência de picos epidêmicos. A terceira e última, de acordo com a teoria inicial de Omran, trata-se da era das doenças degenerativas e causadas pelo homem. Essa etapa é descrita como o momento em que a mortalidade continua diminuindo até se aproximar da estabilidade (OMRAN, 1971). Após revisar sua teoria em 1998, Omran inclui uma quarta fase denominada era do declínio da mortalidade cardiovascular, caracterizada pelo envelhecimento populacional e modificação do estilo de vida, doenças emergentes e doenças antes controladas voltam a aparecer. A quinta fase, dita “futurista” pelo autor, é a era da aspiração pela qualidade de vida e persistência de desigualdades na longevidade (OMRAN, 1998).

Ainda à luz de sua teoria, Omran chama atenção para três modelos básicos de transição epidemiológica, diferenciados de acordo com características socioeconômicas de cada população. O modelo clássico ou ocidental, evidenciado pelas experiências da Inglaterra, País de Gales e Suécia, tendo fatores socioeconômicos como principais determinantes da transição. Outro modelo é o de transição acelerada, exemplificado pelo o que foi observado no Japão. Ocorreu muito mais rápido do que o modelo clássico e foi determinada por avanços sanitários, médicos e sociais. O terceiro modelo é denominado contemporâneo ou tardio descrito pelos casos do Chile e Ceilão (atual Sri Lanka). Esse modelo se encaixa aos padrões observados em países em desenvolvimento, cuja a queda da mortalidade ocorreu de maneira rápida e abrupta pós segunda guerra, muito em função de avanços na saúde pública e importação de tecnologia médica (OMRAN, 1971).

Entretanto, outra linha teórica considera a formulação original e revisões da teoria de Omran insuficientes para tentar explicar a complexidade de vários padrões entre e dentro de diferentes países com modelo de desenvolvimento econômico marcado por desigualdades sociais e, conseqüentemente, condições de vida também desiguais. Além disso, críticas são feitas no sentido de que a transição é definida como intervalos de tempo, lineares e unidirecionais, com início e fim (FRENK et al., 1991a). A ideia dessa outra abordagem é pautada em uma teoria mais abrangente chamada transição de saúde (LERNER, 1973; FRENK et al., 1991a; CALDWELL, 1993), onde a transição epidemiológica é parte dela juntamente com a transição de cuidados de saúde (FRENK et al., 1991a).

Nessa ótica, transição epidemiológica é definida como um processo dinâmico de mudança a longo prazo nas condições de saúde de uma sociedade, incluindo mudanças nos padrões de doença, incapacidade e morte. Essas mudanças se dão em resposta a mudanças demográficas, socioeconômicas, tecnológicas, políticas, culturais e biológicas mais amplas (FRENK et al., 1991a). Essa teoria foi formulada para tentar explicar a transição epidemiológica no contexto de países de renda média. Diante disso, algumas diferenças são identificadas na teoria de Frenk e colaboradores frente à de Omran.

Ao contrário da visão linear e unidirecional de Omran, Frenk et al. (1991a) definem um padrão básico de mudanças no perfil da mortalidade, morbidade e incapacidade. Esse padrão é dado por alterações na composição da mortalidade por causa de morte (redução da participação de doenças infecciosas e aumento das doenças não transmissíveis e lesões); na estrutura etária da mortalidade (em função da transição demográfica e do envelhecimento populacional o peso das mortes e das doenças passam a ser maior nos mais velhos); e no aumento do peso relativo

da morbidade em relação à mortalidade (passando da característica aguda para crônica) (FRENK et al., 1991a).

Porém, há uma heterogeneidade quanto à direção, à sequência dos estágios, e, sobretudo, à progressão diferenciada entre diferentes estratos sociais. Em relação à direção, pode ocorrer o que os autores chamaram de “contra-transição”, ou seja, doenças que já foram controladas podem reaparecer como é o caso da AIDS, malária e dengue, e novas doenças surgirem como a Zika. Além disso, os estágios não são disjuntos e podem ocorrer “sobreposição” dado o ritmo de mudança diferenciado entre estratos da população. Esses diferenciais entre estratos sociais em conjunto com a “sobreposição” persistindo por um longo período dão origem ao que Frenk et al. (1991a) chamaram de modelo de transição epidemiológica “polarizado prolongado”.

Para uma avaliação empírica dessa teoria, Frenk et al. (1991b) selecionaram dois países da América Latina e Caribe: Costa Rica e México. Ao analisar as razões entre as taxas de mortalidade por doenças crônicas e a mortalidade por doenças infecciosas, eles observaram que essa razão era 0,23 em 1930 e passou para 8,05 em 1984 na Costa Rica. No caso do México, essa taxa saiu de 0,26 em 1960 para 1,4 em 1984. Isso significa que, embora os dois sejam países em desenvolvimento, estavam em momentos distintos de suas transições epidemiológicas. A Costa Rica em uma fase mais avançada e o México se encaixa bem no modelo “polarizado prolongado”.

No caso do Brasil, estudos empíricos têm apontado que o país também se enquadra no modelo “polarizado prolongado” descrito por Frenk e seus colaboradores (FRENK et al., 1991b; PRATA, 1992; SCHRAMM et al., 2004; BORGES, 2017). Em um estudo mais recente, Borges (2017) analisou a transição em saúde brasileira a partir das estimativas de expectativa de vida do IBGE para os anos censitários de 1980, 1991, 2000 e 2010 e a contribuição das causas de morte nessas estimativas. Ele conclui que houve uma redução considerável da mortalidade infantil por doenças infecciosas e parasitárias e, posteriormente, por causas relacionadas ao período perinatal, bem como o aumento da participação de doenças crônicas e degenerativas como principal causa de morte. Outro aspecto importante da transição no Brasil é a redução das mortes por doenças cardiovasculares e da persistência de doenças infecciosas e parasitárias (apesar do declínio da mortalidade por essa causa). Entretanto, o autor chama atenção para as mesmas questões levantadas por Frenk et al. (1991b). Ocorrência de “contra-transição”, “sobreposição” e ocorrência de desigualdades regionais e sociais no processo, o que resulta no modelo “polarizado prolongado”.

Nesse contexto de desigualdades regionais e sociais em saúde aliada a mudanças na dinâmica populacional e no padrão epidemiológico, é preciso avançar no conhecimento dos diferenciais de mortalidade adulta segundo o *status* socioeconômico (SSE) por meio da dimensão escolaridade nas grandes regiões brasileiras.

Há tempos a escolaridade vem sendo apontada como uma das mais importantes dimensões do status socioeconômico para estudar diferenciais na mortalidade (KITAGAWA; HAUSER, 1973; PRESTON; TAUBMAN, 1994; MULLER, 2002). Se escolaridade é uma dimensão do *status* socioeconômico e as transições demográfica, da mortalidade e epidemiológica ocorrem de maneira desiguais entre estratos sociais, então espera-se que pessoas menos escolarizadas estejam mais “atrasadas” nesses processos de mudança em relação às mais escolarizadas. Por exemplo, os menos instruídos tendem a morrer mais precocemente por causas externas (homicídio) do que os mais instruídos (FILHO et al., 2007).

Outra questão é que pessoas com mais baixo grau de instrução tendem a ser acometidas por causas de morte que os mais escolarizados já não morrem tanto (causas externas, doenças infecciosas, por exemplo), muito em função de uma maior exposição ao risco, do acesso a serviços de saúde e condições/estilo de vida também desiguais. Embora o Sistema Único de Saúde – SUS tenha contribuído consideravelmente para a redução das desigualdades no acesso aos serviços de saúde, elas ainda persistem (TRAVASSOS et al., 2006; BARATA, 2009).

Travassos et al. (2006) avaliaram essas desigualdades no acesso aos serviços de saúde nas grandes regiões e segundo a renda e escolaridade no período de 1998 e 2003 utilizando dados das PNADs. Os resultados indicaram que o acesso sofre forte influência dessas variáveis. Indivíduos que residiam em regiões com maior grau de desenvolvimento socioeconômico (Sudeste e Sul) tiveram maior acesso aos serviços de saúde comparado aos que moravam em regiões menos desenvolvidas, além de ter se observado uma piora na desigualdade regional entre os dois pontos no tempo. Por outro lado, as desigualdades sociais reduziram no período. Embora a renda tenha apresentado maior influência do que a escolaridade, a chance de um indivíduo com 9 anos ou mais de estudo ter utilizado serviços de saúde no Brasil foi 20,9% maior do que aquele com 0 a 4 anos de escolaridade. Quando analisaram por região, as razões de chance foram estatisticamente diferentes de 1 apenas para as regiões Norte (1,380) e Sudeste (1,281) (TRAVASSOS et al., 2006).

Por isso, esses diferenciais por escolaridade e regiões podem evidenciar as iniquidades em saúde existentes no Brasil por meio da dimensão chamada de situação em saúde (NUNES

et al., 2001). Ademais, estudar a mortalidade por escolaridade também pode ser útil para subsidiar cálculos previdenciários baseados em estimativas de mortalidade diferentes por essa variável.

Além disso, as investigações acerca dessa dimensão são bastante recentes no cenário nacional. Enquanto estudos sobre esses diferenciais ganhavam força nos Estados Unidos na década de 1970 com o estudo pioneiro de Kitagawa e Hauser (1973) evidenciando uma relação inversa entre morte adulta e escolaridade, no Brasil os primeiros estudos iniciaram apenas na década de 2000 (PÉREZ; TURRA, 2008; PÉREZ, 2010; GUEDES et al., 2011; SILVA, 2014; RIBEIRO, 2016; SILVA et al., 2016; TURRA et al., 2016).

Entretanto, esses estudos para o Brasil esbarraram em basicamente dois problemas nos quantitativos de óbito. O primeiro é a má qualidade dos dados de escolaridade do óbito do Sistema de Informação sobre Mortalidade - SIM do Ministério da Saúde/DATASUS sob a dimensão completitude. Em 2010, por exemplo, o percentual de dados faltantes da escolaridade do falecido em idade adulta foi de 29,1%. A estratégia alternativa foi utilizar dados do Censo Demográfico ou da PNAD para estimar de maneira indireta as mortalidades por escolaridade via escolaridade da mãe (PÉREZ; TURRA, 2008; PÉREZ, 2010; GUEDES et al., 2011) e escolaridade da pessoa de referência do domicílio (SILVA, 2014; SILVA et al., 2016). Outra estratégia foi imputar a escolaridade dos óbitos do Censo fazendo uso de variáveis do domicílio e da pessoa de referência numa tentativa de obter a informação individual da pessoa que morreu (RIBEIRO, 2016).

O segundo é a questão da sub-enumeração dos óbitos. Sob essa dimensão da qualidade dos dados de óbito no Brasil, muitos trabalhos têm apontado para uma melhoria da cobertura, porém de forma desigual entre regiões e UFs brasileiras (QUEIROZ, 2012; QUEIROZ et al., 2017). Entretanto, estudos para estimar os fatores de correção por escolaridade, região e sexo utilizando dados do SIM ainda não foram desenvolvidos. Um trabalho que mais se aproximou disso foi o de Silva (2014) e posteriormente Silva et al. (2016). Eles corrigiram os óbitos por escolaridade, mas estimando fatores de correção da população geral para corrigir apenas os óbitos onde a pessoa de referência possuía os menores níveis de escolaridade registrados no censo 2010. A suposição é que a concentração de sub-registros ocorre nos menos instruídos.

Este estudo se insere nessa discussão ao propor uma correção na variável “anos de estudo” dos óbitos adultos registrados em 2010 no SIM/DATASUS através do uso de técnicas de imputação de dados faltantes, assim como uma avaliação de dois cenários (sem correção e

com correção dos óbitos que possuem baixa escolaridade). Esses procedimentos viabilizarão a utilização da informação completa de escolaridade individual do próprio óbito para obter os diferenciais de mortalidade referente a esse atributo, bem como a obtenção de estimativas mais robustas de taxas específicas de mortalidade e probabilidades de morte em cada faixa de escolaridade por região, sexo e idade.

Com isso, será possível chegar a algumas respostas à seguinte pergunta: Quais os diferenciais de mortalidade adulta segundo a escolaridade, sexo e idade no Brasil e grandes regiões em 2010? Nesse contexto e baseado no referencial teórico sobre diferenciais educacionais na mortalidade, algumas hipóteses foram levantadas e serão testadas ao longo do estudo. No que se refere à qualidade dos dados sob a dimensão completitude espera-se que existam diferenciais sociodemográficos nas chances de um registro de óbito apresentar dado faltante na variável escolaridade. Também se acredita que, ao imputar os dados de escolaridade, a distribuição percentual de óbitos por idade e sexo não difiram comparando a distribuição percentual da escolaridade sem imputar. Em relação aos diferenciais por escolaridade, a literatura indica uma relação inversa entre mortalidade e escolaridade, ou seja, quanto maior a escolaridade menor serão as probabilidades de morte. Além disso, os diferenciais de mortalidade por escolaridade são maiores para os homens, diminuem com a idade, a associação negativa entre mortalidade e escolaridade persiste em todas as faixas etárias adultas e a magnitude dos gradientes é diferente entre as grandes regiões brasileiras.

Para elucidar essas respostas, esta dissertação é dividida em seis capítulos. O primeiro se refere a esta introdução. O segundo faz uma revisão da literatura dos diferenciais de mortalidade por escolaridade em vários países do mundo, desde os mais desenvolvidos até os países em desenvolvimento. Além disso, os principais mecanismos explicativos das desigualdades em saúde são revisados. O terceiro capítulo trata de uma avaliação da qualidade da informação sob a dimensão completitude da escolaridade do óbito no SIM em 2010, objetivando identificar se as chances de se observar o registro de escolaridade faltante diferem segundo variáveis sociodemográficas. Esse capítulo será de grande importância para a definição dos estratos utilizados no método de imputação.

O quarto capítulo é dedicado a uma breve revisão da utilização dos métodos de imputação no contexto de estudos sobre diferenciais de mortalidade por nível educacional no Brasil. Nesse mesmo capítulo, são apresentados alguns conceitos importantes sobre imputação e alguns métodos disponíveis, com ênfase na Imputação Múltipla e o método *Hotdeck* com seleção via *ABB – Approximate Bayesian Bootstrap*. A apresentação dos resultados obtidos

com a imputação e as estimativas das variâncias das contagens dos óbitos em cada recorte e seus intervalos de confiança são outros pontos importantes abordado no quarto capítulo. No quinto capítulo são expostas as estratégias de correção de sub-registro, a compatibilização da variável escolaridade do SIM com a do Censo 2010 e o cálculo das medidas de exposição utilizadas nas taxas de mortalidade. Além disso, sobretudo, são apresentados os resultados dos diferenciais de mortalidade por escolaridade por meio das taxas específicas de mortalidade (TEM), a razão de risco (RR), as probabilidades de morte (${}_{35}q_{25}$) e as razões de probabilidade (RP) considerando dois cenários de correção de sub-registro. Por fim, o sexto capítulo apresenta as considerações finais desta dissertação, abordando as relações de seus resultados com o que foi encontrado na literatura, suas limitações e possibilidades de estudos futuros.

1.1 Objetivos

1.1.1 Objetivo geral

Diante do contexto apresentado e visando responder à pergunta de partida da pesquisa, o objetivo geral deste trabalho é estimar os diferenciais da mortalidade adulta no Brasil e regiões usando imputação para compensar dados faltantes nos registros de escolaridade do óbito.

1.1.2 Objetivos específicos

- Avaliar os microdados do SIM/DATASUS quanto a qualidade da declaração de escolaridade e seus determinantes sociodemográficos;
- Imputar os dados ignorados nos registros de escolaridade (variável anos de estudo) dos óbitos registrados no SIM em 2010;
- Construir intervalos de confiança para as contagens de óbitos por idade em cada recorte obtidos através da imputação múltipla;
- Realizar correção de sub registros de óbitos com baixo nível de escolaridade do SIM 2010 para cada recorte;
- Calcular as medidas de exposição com dados do Censo 2010 para cada grupo de escolaridade;
- Estimar as taxas específicas de mortalidade, as razões de risco, as probabilidade de morte entre os 25 e 60 anos(${}_{35}q_{25}$) e as razões de probabilidade (RP)e seus respectivos intervalos de confiança para cada nível de escolaridade segundo sexo, grandes regiões e Brasil;
- Comparar os diferenciais por escolaridade na mortalidade entre os diferentes recortes e entre dois cenários (com e sem correção de sub-registro).

2 DIFERENCIAIS DE MORTALIDADE: REVISÃO DE LITERATURA

Diante desse contexto apresentado na Introdução, faz-se necessário compreender como esses diferenciais já foram documentados em diversos países do mundo. A literatura internacional tem mensurado esses diferenciais na mortalidade baseado em dimensões como renda, ocupação e educação como indicadores do *status* socioeconômico ou de estrato social em uma população.

Entretanto, o nível educacional vem sendo apontado com certas vantagens frente aos demais. Essa dimensão sofre menos mudanças a partir 25 anos de idade, não é influenciado pelas doenças adquiridas no curso da vida adulta e pode ser medida sem que o indivíduo esteja no mercado de trabalho (KITAGAWA; HAUSER, 1973; LIBERATOS et al., 1988; PRESTON; TAUBMAN, 1994; ELO; PRESTON, 1996; MEARA et al., 2008; MONTEZ et al., 2012). Em um estudo para os estados americanos nos anos de 1989 e 1990, o efeito da renda na mortalidade desapareceu quando a variável sobre a escolaridade foi adicionada aos modelos de regressão (MULLER, 2002). Nesse estudo, o autor conclui que a educação é o melhor preditor para explicar variações na mortalidade entre os estados norte-americanos.

Apesar dessas vantagens destacadas para educação, as três dimensões estão totalmente relacionadas entre si. De maneira geral, pessoas com níveis mais altos de educação tendem a ter maiores salários, que por sua vez são vinculados a ocupações privilegiadas, gerando maiores nível de renda para esses indivíduos (PRESTON; TAUBMAN, 1994).

Nos países desenvolvidos existe uma ampla literatura sobre esse tema, já nos países em desenvolvimento, no Brasil por exemplo, a produção é escassa em virtude da dificuldade de obtenção e confiabilidade dos dados para a produção de estimativas robustas (PEREZ, 2010). Enquanto nos países onde não há problemas graves na qualidade dos dados de óbito (EUA, por exemplo) são realizadas estimativas de diferenciais socioeconômico na mortalidade adulta com métodos clássicos e diretos da Demografia, no Brasil é necessário buscar meios de estimá-los de maneira indireta e com vários artifícios metodológicos para contornar a problemática da qualidade dos dados.

Nesse contexto, um ponto de partida de uma revisão de literatura sobre diferenciais socioeconômicos na mortalidade adulta é o estudo clássico e pioneiro de Kitagawa e Hauser (1973). Os autores abordaram as principais dimensões de SSE - *status* socioeconômico (Educação, renda e ocupação) segundo diversas características individuais como sexo, idade, raça, estado civil, natividade, país de origem, parturição, causas de morte e recortes geográficos.

Eles vincularam dados dos registros de óbitos nos Estados Unidos de maio a agosto de 1960 com os dados obtidos no censo demográfico americano do mesmo ano. O objetivo desse pareamento foi obter uma base de dados unificada para que fosse possível analisar diferenciais por características que só estavam disponíveis no censo.

Em relação aos inúmeros resultados observados sobre os diferenciais de mortalidade, mais especificamente por escolaridade, destaca-se o gradiente entre o grupo das mulheres brancas de 25 a 64 anos que possuem graduação ou um nível maior em relação às que possuem menos de 5 anos de estudo. A razão da taxa de mortalidade entre esses dois grupos foi de 2,05. Fazendo essa mesma comparação relativa só que para os homens, a razão foi menor, 1,64. Essas reduções das taxas de mortalidade na medida que aumenta a escolaridade foram observadas em todos os recortes de sexo e idade na população branca americana em 1960 (KITAGAWA; HAUSER, 1973).

A partir de então, vários trabalhos foram desenvolvidos para investigar os diferenciais socioeconômicos na mortalidade utilizando a educação como dimensão relevante para mensurar o status socioeconômico de uma população (ELO; PRESTON, 1996; MONTEZ et al., 2009; MONTEZ et al., 2012; HUMMER; HERNANDEZ, 2013).

Ainda nos Estados Unidos, Elo e Preston (1996) mensuraram os gradientes de mortalidade por escolaridade segundo sexo, idade (25-69 anos; 65-69 anos), compararam com países Europeus e analisaram sob uma perspectiva multivariada. Os autores utilizaram uma base de dados que revolucionou os estudos sobre os efeitos dos diferenciais nas características demográficas e socioeconômicas sobre a mortalidade nos EUA. Trata-se da *National Longitudinal Mortality Study (NLMS)*. Essa base combina dados da pesquisa por amostra de domicílios – a *Current Population Survey (CPS)* – e da *National Death Index (NDI)* base de dados onde são cadastrados os óbitos nos EUA. O estudo de Elo e Preston utilizou dados de 1979-1985.

Os principais resultados de Elo e Preston (1996) indicaram que uma maior escolaridade significa menor risco de morte, em que, exceto para o grupo das mulheres de 25 a 64 anos, os graduados tendem a ter menor mortalidade dos que possuem ensino médio para ambos os sexos e grupos de idade. Outro achado foi que a redução da mortalidade para adultos jovens de 35 a 54 anos quando se incrementa um ano de escolaridade converge para o que é observado por Valkonen (1989) na Europa, porém na Europa a resposta a essa redução é mais sentida nos homens do que nas mulheres (ELO; PRESTON, 1996).

Mais recentemente, Montez et al. (2012) avançaram nessa discussão ao investigar a associação entre o êxito educacional e a mortalidade adulta nos Estados Unidos segundo sexo, idade e raça. Esses autores também utilizaram dados do NLMS, mas no período de 1979-1998. Após vários critérios de exclusão, 1.008.215 brancos e negros não hispânicos entre 25 e 97 anos de idade foram incluídos no estudo, sendo 164.289 (16,3%) desses identificados como falecidos no NDI. O diferencial desse estudo é que os autores realizaram uma comparação de vários modelos de formas funcionais para identificar o que melhor se aplicava para medir essa associação (via *BIC – Bayesian Information Criterion*) em cada recorte.

Entre os resultados obtidos, destacam-se os que indicaram associação negativa entre educação e risco de mortalidade mais acentuada entre as pessoas de 25 a 64 anos do que entre aquelas com 65 anos ou mais dentro de cada grupo de raça e sexo e que o ritmo de redução do risco de morte é mais acelerado após obtenção do ensino médio. Para as mulheres não-hispânicas, de 24 a 64 anos e negras, os riscos decaem exponencialmente após 11 anos de estudo (MONTEZ et al., 2012).

Apesar de Kitagawa e Hauser (1973) terem evidenciado maiores diferenciais entre as mulheres com dados de 1960 para os EUA, estudos empíricos mais atuais demonstram que os diferenciais educacionais na mortalidade geralmente são maiores entre os homens, mas em uma magnitude não tão grande (MONTEZ et al., 2009; HUMMER; LARISCY, 2011; NATIONAL CENTER FOR HEALTH STATISTICS, 2012). De acordo com dados apresentados pelo *National Center for Health Statistics (NCHS)* no relatório de 2011 sobre o estado de saúde da população norte-americana, a diferença entre a expectativa de vida aos 25 anos dos americanos não institucionalizados do sexo masculino com maior nível de escolaridade e menor nível de escolaridade foi de 9,3 anos, enquanto que para mulheres essa diferença foi de 8,6 no ano de 2006 (NATIONAL CENTER FOR HEALTH STATISTICS, 2012).

Diferenças entre sexos também foram evidenciadas (para homens e mulheres brancos não-hispânicos com 55 anos ou mais no período 1986-2002) em termos dos riscos de morte estimados por meio de modelos de riscos proporcionais de Cox, sendo que as diferenças foram atribuídas aos homens solteiros (MONTEZ et al., 2009). Uma hipótese lançada pelos autores para explicar esses diferenciais é que os comportamentos de saúde e estado de saúde dos homens com baixo nível socioeconômico são particularmente vulneráveis à falta de laços sociais que melhoram a saúde, como o casamento. Outra explicação é que a educação como capital humano pode ter efeito maior nos homens por ser capaz de mudar decisões de

comportamentos de riscos, principalmente aqueles relacionados à mortes por causas externas (homicídio e acidente de trânsito, por exemplo) (ROSS et al., 2012).

Porém, Zajacova (2006) chama atenção para resultados controversos na literatura sobre essas diferenças de gênero, alegando que há trabalhos em que os diferenciais são maiores para os homens, outros para as mulheres e tem aqueles que não identificam diferenças. No trabalho da autora, a escolaridade teve o mesmo efeito na mortalidade para homens e mulheres. Cada ano de escolaridade adicional foi associado a uma menor chance de morrer em torno de 5% para ambos os sexos. Outros trabalhos apresentaram conclusões que corroboram a ideia de que, geralmente, o sucesso educacional não é diferenciado por sexo (ZAJACOVA; HUMMER, 2009; HUMMER; LARISCY, 2011). Ao contrário dos EUA, maiores desigualdades entre os homens parecem ser mais claras na Europa (MACKENBACH et al., 2008) e em países em desenvolvimento da América Latina (SANDOVAL; TURRA, 2015; SILVA et al., 2016). Provavelmente, esse desalinhamento entre os resultados ocorra em função da utilização de diferentes metodologias.

Em relação ao comportamento dos gradientes por idade no contexto americano, desde o estudo de Kitagawa e Hauser eles eram maiores entre os adultos mais jovens, ou seja, as diferenças reduziam com a idade (HUMMER; HERNANDEZ, 2013). No estudo de Hummer e Lariscy (2011), a razão entre o risco de morte em indivíduos com 12 anos de estudo e o risco em indivíduos com mais de 12 anos de estudo é maior nas idades de 25 a 44 anos. Essas diferenças relativas diminuem nos grupos de idade subsequentes (45 a 64 anos e 65 a 84 anos) para os dois sexos. Uma das razões mais prováveis indicada na literatura é a seletividade da mortalidade (CRIMMINS, 2005).

Diante do que foi apresentado, percebe-se que a relação inversa entre educação e mortalidade já está devidamente documentada no contexto norte-americano. Porém, as investigações também avançaram nos EUA no sentido de verificar como esses diferenciais se comportaram ao longo do tempo dado a queda observada na mortalidade como um todo. As desigualdades em saúde diminuíram nos EUA segundo o status sócio econômico (SSE)? Em outras palavras, a queda da mortalidade foi observada igualmente para todos os estratos sociais? Estudos empíricos mostraram que não.

Nesse contexto de queda na mortalidade geral nos EUA, com grande influência da redução sustentada da mortalidade por doenças cardiovasculares (STAMLER, 1985), Pappas et al. (1993) compararam seus resultados para o ano de 1986 com os obtidos por Kitagawa e

Hauser (1973) para o ano de 1960. Embora tenha sido utilizado apenas dois pontos no tempo para realizar essa comparação e oscilações nos diferenciais de mortalidade por SSE possam ter ocorrido dentro desse intervalo temporal, os autores mantiveram os critérios metodológicos similares aos utilizados por Kitagawa e Hauser (1973) para que fosse possível realizá-la. Por exemplo, eles calcularam as taxas de mortalidade padronizadas para o grupo de indivíduos adultos de 25 a 64 anos por raça, sexo, renda e nível educacional.

Assim, evidenciou-se que a relação inversa entre mortalidade e SSE foi mantida ao longo do tempo, porém as desigualdades nas taxas de mortalidade foram ampliadas. Para o grupo das mulheres (tanto brancas quanto negras), os gradientes segundo a escolaridade aumentaram mais de 20%, enquanto que nos homens (tanto brancos quanto negros) esse aumento foi de mais de 100%. Vale ressaltar que, embora as taxas de mortalidade tenham diminuído para todos os níveis de escolaridade, a redução foi sentida com mais força para homens e mulheres com maiores níveis educacionais. (PAPPAS et al., 1993). Evidências dessa tendência de piora na desigualdade da mortalidade nos EUA também foram observadas ao comparar os anos de 1981-1988 e 1991-1998 com dados do *National Longitudinal Mortality Study - NLMS* e 1990-2000 da *Multiple Cause of Death – MCD* combinada com dados do censo (MEARA et al., 2008). Montez e colegas também chegaram a essas conclusões utilizando dados combinados do *National Health Interview Survey – NHIS* com os registros de óbito do *National Death Index – NDI* para os períodos de 1986-1992, 1993-1999 e 2000-2006 (MONTEZ et al., 2011).

Possíveis explicações para o aumento dessa desigualdade socioeconômica na saúde e mortalidade são dadas no sentido de que parte da população americana tenha mudado seu estilo de vida no período analisado. Diante de uma forte redução da mortalidade por doenças cardíacas coronárias, fatores como mudanças nos hábitos alimentares com a ingestão de alimentos mais saudáveis, redução das taxas de prevalência do tabagismo na população adulta e aumento da prática de atividade física são colocados como fundamentais para redução dessa causa de morte. Contudo, essa mudança de hábitos ocorreu mais fortemente entre os indivíduos mais bem instruídos comparado aos menos instruídos (STAMLER, 1985). No mesmo sentido, Meara et al. (2008) mostraram que 20% da tendência observada, ou seja, do aumento dos gradientes na mortalidade no ano de 2000 comparado à 1990, são devido às tendências também seletivas por escolaridade observadas nas doenças relacionadas ao tabagismo como câncer de pulmão e doença pulmonar obstrutiva crônica.

A relação inversa entre SSE e mortalidade adulta também vem sendo bem documentada em países europeus. Alguns pesquisadores, como por exemplo Tapani Valkonen, Anton Kunst e Johan Mackenbach, tiveram papel importante para manter uma agenda de pesquisa sobre essa temática na Europa. Além de mensurar as desigualdades em saúde através dos diferenciais por SSE, eles também realizam estudos para monitorar esses gradientes ao longo do tempo em diversos países europeus (VALKONEN, 1989; KUNST; MACKENBACH, 1994; MACKENBACH et al., 1997; KUNST et al., 2004; MACKENBACH et al., 2008; MACKENBACH, 2012; MACKENBACH et al., 2015; MACKENBACH et al., 2016; MACKENBACH et al., 2018).

De maneira singular, Mackenbach et al. (2008) reuniram dados de pesquisas realizadas em 22 países europeus com o objetivo de comparar a magnitude das desigualdades em saúde por meio de diferenciais educacionais, ocupacionais e de renda na mortalidade por idade (30 a 74 anos), sexo e causa de morte com períodos variando do início da década 1990 ao início da década 2000. Diferente de estudos anteriores (KUNST; MACKENBACH, 1994; MACKENBACH et al., 1997), nesse eles ampliaram o leque ao incluir países da Europa Oriental. Em relação à dimensão escolaridade, eles utilizaram a variável categorizada como até 6 anos de estudo, de 7 a 9, 10 a 11 e 12 ou mais anos de estudo disponíveis com os mesmos níveis em quase todos os países, o que torna os dados comparáveis. Para estimar as taxas de mortalidade, eles utilizaram modelo de regressão de Poisson e apresentaram os resultados por meio do IRD - Índice Relativo de Desigualdade (razão entre a taxa de mortalidade no grupo de menor e maior escolaridade).

Os resultados indicaram que, para todos os países analisados e ambos os sexos, o IRD foi maior que 1. Isso significa que o risco de morte entre os menos escolarizados é maior do que os mais escolarizados. A magnitude desses gradientes é diferente entre os países e dentro dos países por sexo. Entre os homens, as maiores desigualdades educacionais na mortalidade foram observadas na República Tcheca, Hungria, Polônia e Lituânia e as menores em países como Suécia, Noruega, Finlândia, Suíça e Bélgica. Nas mulheres esse padrão se repete, mas em menor magnitude em alguns países. Maiores gradientes entre os homens são observados na maioria dos países analisados, conforme observado na literatura para os EUA. Porém esse padrão não é generalizado. Por exemplo, na República Tcheca as desigualdades de gênero na mortalidade por nível educacional são evidentes. O IRD para os homens ficou em torno de 4,5 enquanto para as mulheres esse índice foi de 2,5, indicando um maior diferencial entre os homens. Por outro lado, na Noruega e Dinamarca, por exemplo, os índices ficaram em torno de

2,0 tanto nos homens quanto nas mulheres, o que indica não haver desigualdade de gênero nas mortes segundo a escolaridade do indivíduo (MACKENBACH et al., 2008).

Em relação aos gradientes por idade, há evidências de que eles diminuem com a idade em alguns países europeus. Para o período de 1990-1994, Kunst et al. (2004) mostraram que na Noruega o IRD saiu de 3,85 na idade de 30 a 44 anos para 2,48 na faixa de 45 a 59 anos e para 1,70 nos idosos de 60 a 74 anos. Na Finlândia, esse gradiente foi de 3,36, 2,22 e 1,80.

Um caso que chama atenção nos estudos dos diferenciais socioeconômicos na mortalidade na Europa e que não foi incluído no estudo de Mackenbach et al. (2008) é o da Alemanha. Curiosamente, esse país não dispõe de um registro nacional de mortalidade nem de pesquisas que acompanhem os indivíduos inquiridos em censos como é observado em outros países europeus (KROLL et al., 2017). Uma tentativa de estimar esses diferenciais por educação, ocupação e renda foi realizada com base em uma combinação da pesquisa *German National Health Interview and Examination Survey* realizada em 1998 com as tabelas de sobrevivência de período do *German Statistical Office* por meio de modelos paramétricos de sobrevivência de Gompertz. Os resultados foram consonantes ao que é observado na literatura norte-americana e europeia, onde a expectativa de vida ao nascer no grupo de homens de baixo nível socioeconômico foi de 71 anos contra 80 anos dos que possuem alto nível socioeconômico. Para as mulheres esses valores são de 78 e 86 anos, respectivamente (KROLL et al., 2017). Assim como nos EUA, na Europa essas desigualdades socioeconômicas na mortalidade e em saúde também persistem ao longo do tempo. As desigualdades relativas na mortalidade em alguns países europeus tenderam a aumentar durante as últimas décadas do século XX (KUNST et al., 2004). Entretanto, e apesar da crise financeira de 2008, há evidências de uma redução em termos absoluto na desigualdade para alguns países entre os períodos de 1990 e 2010, provavelmente em função de mudanças comportamentais e melhorias na prevenção e tratamento da saúde (MACKENBACH et al., 2016; MACKENBACH et al., 2018). A ampliação das desigualdades em períodos mais recentes parece ser uma particularidade norte-americana no contexto de países desenvolvidos (BOSWORTH, 2018). Mesmo em uma situação bastante favorável em termos de disponibilidade e qualidade dos dados (como é o caso dos países europeus), os resultados de pesquisas que comparam mudanças na desigualdade em saúde via mortalidade entre países devem ser vistos com parcimônia devido à utilização de diferentes fontes de dados e critérios distintos utilizados nas pesquisas dos diferentes países (KUNST et al., 2004). Ainda assim, os resultados são consistentes em evidenciar uma relação inversa entre morte e escolaridade na Europa. Uma revisão das teorias que podem ajudar a

explicar a manutenção dessas diferenças (o autor denomina como um “paradoxo”) no contexto europeu são discutidas e podem ser encontradas em (MACKENBACH, 2012).

Em países em desenvolvimento esses estudos são mais recentes e escassos comparados aos realizados nos países desenvolvidos. Isso se deve em grande parte à indisponibilidade de pesquisas apropriadas para estudos sobre diferenciais educacionais na mortalidade e da má qualidade dos dados de óbito por essa variável naqueles países. Nesse contexto, podem-se destacar os estudos realizados para a China (LOWRY; XIE, 2009), Índia (SAIKIA; BHAT, 2008), em países da América Latina, no Chile (SANDOVAL; TURRA, 2015), na Argentina (MANZELLI, 2014) e no Brasil (PÉREZ; TURRA, 2008; PÉREZ, 2010; GUEDES et al., 2011; SILVA, 2014; RIBEIRO, 2016; SILVA et al., 2016).

Por exemplo, na Índia, pesquisas sobre mortalidade adulta parecem ser escassas devido à falta de dados que permitam tal investigação (SAIKIA; BHAT, 2008). Por ser um país de particularidades culturais e sociais, algumas dimensões de status socioeconômico utilizadas para investigar os gradientes na mortalidade são diferentes das que são utilizadas nos países em desenvolvimento. Alguns pesquisadores levaram em conta, por exemplo, a casta (SUBRAMANIAN et al., 2006; SAIKIA; BHAT, 2008; PO; SUBRAMANIAN, 2011) e padrão de vida (SUBRAMANIAN et al., 2006; SAIKIA; BHAT, 2008; PO; SUBRAMANIAN, 2011). Porém, o estudo de Saikia e Bhat (2008) investigou os aspectos socioeconômicos, demográficos e comportamentais que influenciam a mortalidade de indivíduos indianos de 15 a 59 anos, entre eles o nível de alfabetização do agregado familiar. Apesar de as informações não serem individuais dos óbitos, observou-se menores chances de morte à medida que aumenta o nível de alfabetização no domicílio para ambos os sexos.

Tomando os casos do Chile, Argentina e Brasil para representar o contexto de países em desenvolvimento na América Latina, observa-se que a relação inversa entre mortalidade e escolaridade também foram evidenciadas nesses países. Para Sandoval e Turra (2015), embora haja poucos trabalhos sobre os diferenciais de mortalidade por nível educacional no Chile, os dados de estatísticas vitais (mortalidade) e variáveis sociodemográficas são considerados de boa qualidade, tanto em relação ao grau de cobertura dos registros de óbitos, quanto à completitude da variável escolaridade (percentual de registro ignorado inferior a 1%). Isso permite que estimativas dos diferenciais de mortalidade por escolaridade sejam realizadas sem a necessidade de correções nos dados. Para os períodos de 1991-1993 e 2001-2003 e adultos de 30 anos ou mais, os autores estimaram as taxas de mortalidade por meio de vários modelos de regressão Poisson. Esses modelos tinham como variável resposta o número de mortes e

variáveis explicativas idade, escolaridade, período e uma variável que mensura tempo de exposição (pessoas-ano calculado com dados do Censo 1992 e 2002). Variações de modelos foram elaboradas, onde modelos com efeitos principais e outros com interações entre as variáveis idade e período com educação foram estimados.

Os principais resultados indicam que de maneira geral a relação inversa entre morte e escolaridade foi observada, que os diferenciais foram maiores entre os homens e que diminuem com a idade. Outra constatação foi o aumento das desigualdades ao comparar os dois triênios, onde observaram que houve aumento das taxas de mortalidade entre os menos escolarizados e redução entre os mais escolarizados. Isso pode retratar uma desigualdade nas condições e acesso à saúde diferenciados no Chile (SANDOVAL; TURRA, 2015). Esses resultados vão de encontro ao que foi observado em países desenvolvidos conforme apresentado anteriormente.

Além disso, esses resultados para o Chile se assemelham aos que foram encontrados na Argentina. Entretanto, diferentes metodologias foram empregadas (MANZELLI, 2014). O autor do estudo para Argentina calculou taxas de mortalidade por escolaridade, idade, sexo e região para estudar os diferenciais educacionais na mortalidade em 2010 para óbitos de 25 a 64 anos. O numerador das taxas é o número de óbitos em cada recorte obtidos das estatísticas vitais coletadas por meio da declaração de óbitos. No denominador, medidas de exposição foram calculadas com base em dados do censo. No que se refere à qualidade dos dados, não foi necessário realizar correção de sub-registro, pois o autor destaca que há uma cobertura de 100% na Argentina. Por outro lado, a variável escolaridade apresentou 40% de dados faltantes. Manzelli obteve uma base de dados completa por meio da Imputação Múltipla. Ainda em termos metodológico, análises comparativas merecem cautela, pois não fica claro se foi realizado padronização das taxas de mortalidade, medidas essas influenciadas pela estrutura etária da região.

Ainda assim, os resultados também convergem para o que é observado na literatura: relação inversa entre mortalidade e escolaridade, diferenciais mais expressivos entre os homens e decrescem com a idade. Em relação aos diferenciais regionais da educação na mortalidade, a hipótese inicial dos autores de que em regiões menos escolarizadas os diferenciais seriam maiores foi totalmente refutada. Regiões mais escolarizadas como a de Buenos Aires e Patagônia apresentaram os maiores gradientes (MANZELLI, 2014).

As investigações sobre diferenciais educacionais na mortalidade no Brasil têm avançado à medida que os sistemas de informação em saúde vão avançando. O interesse pelo tema no

Brasil partiu de investigações da influência da escolaridade da mãe na mortalidade infantil. Por exemplo, Fernandes (1984) analisou a mortalidade das crianças em relação à educação da mãe para as regiões metropolitanas do Brasil e os estados onde elas estão contidas e concluiu que as diferenças das estimativas de esperança de vida entre as classes extremas de anos de estudo são altas.

É a partir da década de 2000 que estudos sobre diferenciais educacionais na mortalidade adulta surgem no Brasil. Em um contexto de indisponibilidade de pesquisas apropriadas e de má qualidade dos dados, várias estratégias de análise indireta têm sido adotadas (PÉREZ; TURRA, 2008; PÉREZ, 2010; GUEDES et al., 2011; SILVA, 2014; RIBEIRO, 2016; SILVA et al., 2016).

Pérez e Turra (2008) contribuíram com a análise de dados individuais de mortalidade adulta de mulheres no Brasil segundo o nível de escolaridade da mãe utilizando dados da PNAD de 1996, tendo como um dos principais resultados a constatação de diferença de mortalidade entre mulheres de 20 a 29 anos sem escolaridade e com 9 anos de estudo, numa ordem de 7 vezes maior no grupo sem escolaridade.

Mais recentemente, Silva et al. (2016) estimaram os diferenciais de mortalidade em cada faixa de escolaridade levando em consideração o sexo, idade e recortes geográficos (Brasil e grandes regiões) utilizando dados do censo demográfico de 2010, combinaram a informação do nível de escolaridade do chefe do domicílio com as informações de sexo e idade da pessoa falecida na residência (sob a suposição de que o regime de mortalidade não difere significativamente entre indivíduos de um mesmo domicílio). As medidas de mortalidade utilizadas foram as probabilidades de morte entre 15 e 60 anos ($45q_{15}$) e a expectativa de vida aos 15 anos em cada recorte obtidos por meio das funções da tabela de sobrevivência usual.

Eles concluíram que os diferenciais de mortalidade por escolaridade mais consideráveis foram na população masculina, uma vez que a expectativa de vida de homens com ensino superior é 6,27 vezes maior do que os que possuem um nível abaixo do fundamental completo. Sob o aspecto geográfico/regional Silva et al. (2016) também deram contribuição. Constataram que nas regiões nordeste e sudeste do Brasil esses gradientes de mortalidade por escolaridade são mais evidentes, tanto para o sexo masculino quanto para o feminino.

Ainda no contexto brasileiro, Ribeiro (2016) realizou uma análise dos diferenciais de mortalidade adulta por escolaridade para o estado de São Paulo e sua capital utilizando os dados do censo 2010 por meio da regressão de Poisson. Antes de estimar as taxas de mortalidade por

idade, sexo e escolaridade, foi necessário imputar os dados de escolaridade do falecido, já que os únicos quesitos inquiridos sobre o óbito são sexo e idade (RIBEIRO, 2016). Os principais resultados indicaram um maior diferencial entre os homens (conforme também encontrado nos estudos citados nessa revisão) e na capital São Paulo. Ela também chama atenção para um acentuado diferencial entre a categoria de menor escolaridade e as demais, porém o gradiente entre a média escolaridade e a escolaridade mais alta foi mínima entre os homens e inexistente entre as mulheres do estado de São Paulo. Para a capital, também foram observados maiores diferenciais entre os homens, apresentando uma taxa 4 vezes maior nos menos escolarizados em relação aos mais escolarizados. Enquanto que nas mulheres a taxa foi 2,7 vezes maior.

O único trabalho que fez uso de dados de escolaridade individuais do óbito oriundos do SIM para estimar diferenciais de mortalidade foi o de Turra et al. (2018). Eles utilizaram dados de óbito adulto (45 anos em diante) do SIM de 2014 considerando apenas as capitais que apresentaram um percentual de escolaridade faltante de até 20%. Em seguida agruparam essas capitais por região para estimar os diferenciais por escolaridade segundo sexo, idade e região.

Os principais resultados encontrados por Turra et al. (2018) foram que os diferenciais entre o risco de morte de quem tem até sete anos de estudo em relação aos óbitos com doze anos ou mais de estudo tendem a diminuir com a idade. Esses gradientes são maiores para os homens e também nas capitais das regiões Sudeste e Centro-oeste.

Dado que há uma ampla literatura evidenciando uma relação negativa entre mortalidade e status socioeconômico, principalmente em países desenvolvidos a partir da segunda metade do século XX e que há uma forte relação entre as dimensões de SSE (escolaridade, renda e ocupação), é fundamental compreender e apresentar as teorias que tentam explicar essas desigualdades em saúde. As principais abordagens explicativas dos porquês desses diferenciais são a econômica e a psicossocial (PRESTON; TAUBMAN, 1994).

Considerado por esses autores como o mais sistemático, o modelo econômico funda-se na ideia de que os indivíduos tomam decisões que ajudam a determinar seu estoque de saúde (H) em qualquer período de tempo. Esses indivíduos chegam a óbito quando H declina para níveis abaixo de algum ponto crítico (H^*). O controle desse estoque de saúde não é realizado pelos indivíduos de maneira direta visando estabelecer uma idade para a morte. Esse controle ocorre por meio de decisões de investimentos individuais, por exemplo, em exercícios físicos e hábitos saudáveis de consumo que afetam positivamente o estoque e pode postergar a idade de morrer (PRESTON; TAUBMAN, 1994).

Um pressuposto dessa teoria é que o indivíduo maximiza uma função de utilidade. Essa função depende de variáveis que denotam os itens consumidos pelos indivíduos, sendo esses itens endógenos (relacionados totalmente ou parcialmente ao estoque de saúde) ou exógenos (sem relação direta com a saúde). A alocação de recursos entre esses bens de consumo é dada por duas equações: uma orçamentária e outra de produção de saúde. A orçamentária leva em consideração preço e quantidade, renda do indivíduo, horas trabalhadas e poupança. Enquanto que a de produção de saúde relaciona quantidade de cada item de consumo relacionado à saúde, a quantidade de conhecimento médico, a disponibilidade de instalações médicas, a genética do indivíduo e características ambientais que eles vivenciam (PRESTON; TAUBMAN, 1994).

Aqui é importante destacar como algumas dessas variáveis agem sobre a saúde. Indivíduos com maior escolaridade e maiores rendas possuem maiores possibilidades de adquirir bens e serviços que auxiliem na promoção à saúde e prevenção de riscos (planos de saúde, alimentos saudáveis, por exemplo), bem como de viverem em áreas melhores assistidas pelos serviços públicos. Nesse mesmo sentido, maior escolaridade contribui para um maior grau de conhecimento sobre a medicina e o acesso a tecnologias médicas e, conseqüentemente, em uma melhor saúde e menor mortalidade (PRESTON; TAUBMAN, 1994).

A capacidade de mensurar os efeitos de variáveis endógenas e exógenas nas condições de saúde é apontada por Preston e Taubman (1994) como principal vantagem da teoria economicista. Entretanto, eles chamam atenção para limitações dessa teoria, destacando-se o fato de que o estoque de saúde e renda (a renda com relação forte com a escolaridade conforme foi abordada) possuem uma relação de mão dupla, uma influenciando a outra simultaneamente. Uma alternativa apontada pelos autores para verificar essas causalidades são pesquisas longitudinais com acompanhamento dos mesmos indivíduos. Outra lacuna dessa abordagem é não ter a capacidade de captar o comportamento e as relações interpessoais dos indivíduos.

No contexto de países em desenvolvimento, Caldwell (1990) argumenta que, ao contrário da ideia de Preston e Taubman (1994), características sociais (como o nível de escolaridade ou controle da fecundidade) ou características culturais (como grupo étnico) geralmente são mais influentes na determinação dos níveis de mortalidade do que o acesso a serviços médicos e renda.

É nessa linha de raciocínio que a abordagem psicossocial tenta preencher as lacunas explicativas deixadas pelas teorias economicistas. Fatores psicossociais que podem influenciar a relação entre saúde e status socioeconômico como gostos, satisfações ao longo da vida e

questões relacionadas à personalidade do indivíduo não são levados em consideração pelos economistas (PRESTON; TAUBMAN, 1994). Em contrapartida, apenas fatores dessa natureza são considerados, enquanto as principais dimensões de SSE (educação, renda, ocupação) são preteridos totalmente ou, no máximo, usados como variáveis de controle no contexto psicossocial (COHEN; SYME, 1985; PRESTON; TAUBMAN, 1994). Assim, é importante apresentar os pontos de vista dos defensores da teoria psicossocial como mecanismo explicativo da relação entre SSE e saúde (HOUSE et al., 1988; WILLIAMS, 1990).

No contexto norte-americano, Williams (1990) chama atenção para o fato de que as desigualdades segundo SSE na morbimortalidade foram mantidas ou até mesmo ampliadas, apesar de avanços importantes em alguns fatores que são influenciados pelo mecanismo economicista. Ele argumenta que os bens e serviços que auxiliam na promoção à saúde e prevenção de riscos estão ao alcance da maioria das famílias americanas (nutrição adequada, habitação, água e eliminação de resíduos). Outro ponto importante é a criação do Medicare e Medicaid. Mesmo com a disponibilidade de tratamento médico aos menos favorecidos, as desigualdades de SSE na morbimortalidade não se extinguiram, rebatendo um dos argumentos Preston e Taubman (1994). Isso significa que a teoria economicista é limitada para explicar a associação entre SSE e saúde (WILLIAMS, 1990).

A ideia da corrente psicossocial é de que fatores sociais e psicológicos (comportamentos de saúde, estresse em ambientes familiares, residenciais e ocupacionais, integração social e apoio, e percepções de domínio e controle) são fundamentais na explicação das disparidades socioeconômicas em saúde. Esses fatores são vistos como características do estilo de vida e condições de vida resultantes de respostas dos grupos sociais às realidades e restrições do ambiente externo (WILLIAMS, 1990).

O modelo conceitual apresentado por Williams (1990) propõe que, o SSE possui relação direta como os fatores psicossociais e, por sua vez, têm efeitos diretos nos resultados de saúde. Além disso, esses fatores psicossociais em conjunto com os cuidados médicos interagem com o SSE e exercem efeito conjunto na saúde. O modelo também aborda uma relação de mão dupla entre os fatores psicossociais e os cuidados médicos. Fatores demográficos e genéticos são tratados como variáveis de controle. Percebe-se, portanto, que o modelo apresenta causas complexas para os resultados em saúde.

House et al. (1988) argumentam que estudos empíricos têm evidenciado que indivíduos com menos interação social possuem um maior risco de morte em função de uma saúde

psicológica e física mais fragilizada em relação aos mais socializados. House e seus colaboradores acrescentam que as relações sociais devem ser consideradas no mesmo rol de fatores de risco para a saúde como o tabagismo, pressão arterial, obesidade e atividade física.

Entre as críticas a essa abordagem destacam-se as feita por Preston e Taubman (1994). Por exemplo, quando variáveis que mensuram a personalidade são utilizadas como preditoras em conjunto com a educação para explicar a condição de saúde, modelos de regressão múltipla tradicionais não são eficientes para identificar causalidade dado a multicolinearidade existente entre essas variáveis explicativas. Eles também chamam atenção para esse mesmo problema ligado a fatores de relacionamento social. Além de melhorar a saúde, essas redes sociais também afetam o sucesso profissional do indivíduo e, conseqüentemente, o SSE. Porém, eles reconhecem que é difícil separar o nexos causal entre os fatores psicossociais, status socioeconômico e saúde, se estabelecendo, portanto, uma relação de mão dupla entre os três.

Conforme as delimitações desta dissertação e diante do que foi observado acerca dos diferenciais de mortalidade por escolaridade na literatura nacional e internacional, é razoável esperar uma relação inversa entre o risco de morte e o nível educacional, bem como esses gradientes sejam maiores entre os homens e que reduzam com a idade. Ademais, acredita-se que os gradientes na mortalidade por educação sejam diferentes entre as grandes regiões do Brasil.

Após a contextualização demográfica e epidemiológica, bem como de uma revisão das evidências sobre a relação entre o sucesso educacional e a mortalidade, o próximo capítulo trata sobre a qualidade dos dados de óbitos no Brasil e uma avaliação da dimensão completude da variável escolaridade do SIM é realizada.

3 AVALIAÇÃO DA QUALIDADE DOS DADOS: COMPLETITUDE

Há tempos a qualidade dos dados vem sendo objeto de discussão nas mais diversas áreas do conhecimento. Para AbouZhar e Boerma (2005), a tomada de decisões acertada na área de saúde pública só é possível com informações de qualidade sobre eventos relacionados, dentre outros, à mortalidade e com a existência de um sistema de informação adequado. Dessa maneira, torna-se imprescindível a avaliação da qualidade desses dados antes de prosseguir com análise de diferenças na mortalidade, sob o risco de se tomar decisões errôneas por meio de interpretações inadequadas.

De acordo com Lima et al. (2009), a qualidade de um sistema de informação de saúde vem sendo avaliada pela literatura de maneira multidimensional. Duas dessas dimensões vêm sendo amplamente discutidas no Brasil, a cobertura e a completitude, principalmente em estudos sobre mortalidade, que têm como principal fonte o Sistema de Informação sobre Mortalidade – SIM do Ministério da Saúde. No que diz respeito à cobertura, em função do desenvolvimento e aprimoramento de técnicas demográficas que mensuram sub-registros, a literatura tem avançado bastante no que diz respeito à avaliação do grau de cobertura dos óbitos (PAES; ALBUQUERQUE, 1999; PAES, 2005; QUEIROZ, 2012; QUEIROZ et al., 2017). Essa dimensão da qualidade dos dados é discutida no Capítulo 5.

Em relação à completitude da escolaridade do falecido, se faz necessário definir o conceito de dados ignorados ou faltantes, pois serão utilizados para mensurar a completitude. Rubin (1987) parte de um conceito mais amplo que ele chama de não-resposta. Ele define não-resposta como qualquer situação em que haja valores faltantes na matriz de dados a serem analisados, seja pela recusa de um inquirido responder à pesquisa ou da ausência de resposta para um dos quesitos pesquisados. Aqui, dado ignorado ou faltante é aquele cujo registro foi realizado com a codificação “Ignorada” conforme sugerido pelos manuais de preenchimento da declaração de óbito, ou que foi deixado em branco ou com codificação inexistente pelo responsável do preenchimento ou digitação (ROMERO; CUNHA, 2006; COSTA; FRIAS, 2011; CORREIA et al., 2014; GALDINO et al., 2017).

Ao realizar uma revisão sistemática das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de saúde, após definir alguns critérios de busca, Lima et al. (2009) identificaram 15 estudos que avaliam a completitude envolvendo alguma variável do SIM. Todos abordaram um público alvo e causas de morte específicas. Romero e Cunha (2006), por exemplo, analisaram a completitude de quesitos socioeconômicos e demográficos das

declarações de óbito de crianças menores de um ano no período de 1996 a 2001, segundo regiões e unidades da federação. Uma das conclusões desses autores foi que a escolaridade materna foi classificada como muito ruim (50% ou mais de dados faltantes) na maioria das regiões brasileiras.

Já Felix et al. (2012) avaliaram a completitude das variáveis sexo, idade, raça/cor, escolaridade e estado civil dos óbitos femininos por câncer de mama no período de 1998 a 2007 no Espírito Santo, Sudeste e Brasil registrados no SIM. Eles realizaram uma análise descritiva, de classificação segundo escala de qualidade sugerida por Romero e Cunha (2006) e temporal dos casos ignorados dessas variáveis. Eles concluíram que as variáveis sexo e idade são consideradas excelente. Já as variáveis raça/cor e escolaridade tendem para um percentual de não completitude decrescente no Sudeste e no Brasil. Mais especificamente sobre a incompletitude da escolaridade, Felix et al. (2012) verificaram também elevados níveis de dados ignorados de 1998 a 2007 no Espírito Santo, com uma média de 55,0%, sendo o pior resultado em 1998 (79,0%) e o melhor em 2000 (42,0%), oscilando entre as classificações Ruim e Muito Ruim.

Realizando análises da mesma natureza só que para óbitos por suicídio no Brasil, Sudeste e Espírito Santo de 1996 a 2007, Macente e Zandonade (2010) evidenciaram tendências decrescentes para a incompletitude da variável escolaridade no Brasil e região Sudeste. Elas apresentaram também os percentuais de dados ignorados de escolaridade dos óbitos por suicídio para o Espírito Santo no período. De 1996 a 2007, observaram uma média de 76,1% de dados de escolaridade ignorados, sendo todos os anos classificados como Muito Ruim.

Na mesma temática, mais recentemente, Galdino et al. (2017) avaliaram espacialmente e temporalmente a completitude do quesito “acidente de trabalho” preenchido apenas para os óbitos que tiveram mortes por causas externas para o período de 2007 a 2012. Concluíram que apesar de apresentar melhoria ao longo do tempo, o registro desse quesito é classificado como ruim, chegando a um percentual de 79,1% de não preenchimento ou preenchido com “Ignorado”.

Diante da problemática da falta de qualidade, mais especificamente de completitude, alguns trabalhos têm descartado o uso do SIM por conta do elevado grau de dados ignorados em algumas variáveis da declaração de óbito, muito provavelmente fruto da negligência e preenchimento de maneira inadequada por parte dos profissionais da saúde (NUNES et al., 2001) . Entre aqueles, destacam-se os poucos estudos que objetivaram realizar estimativas de

mortalidade por nível de escolaridade, os quais passaram a utilizar o censo demográfico em detrimento do SIM pelo fato de a variável escolaridade apresentar elevado grau de incompletitude. Por exemplo, Silva et al. (2016) combinaram o novo quesito de mortalidade domiciliar com as informações do nível de escolaridade da pessoa de referência do domicílio utilizando o Censo Demográfico 2010.

Em 2010, 31,96% e 29,14% dos óbitos registrados no SIM considerando todas as idades e para os adultos (desconsiderando os dados faltantes da idade) para o Brasil, respectivamente, tiveram o registro da escolaridade ignorado, sendo classificado como ruim (20% a 50% de ignorados) conforme escala sugerida por Romero e Cunha (2006). A maioria dos trabalhos encontrados na literatura apenas mensuram de maneira descritiva e/ou temporal os dados ignorados das variáveis do SIM, mas não procura relacionar essa incompletitude com outras variáveis, por exemplo, se há seletividade segundo variáveis sociodemográficas. Logo, partindo do fato de que há um percentual considerável de dados faltantes na variável escolaridade do óbito no Brasil em 2010, surge o seguinte questionamento: as chances de a escolaridade do óbito adulto não ser registrada diferem segundo variáveis sociodemográficas? É sob essa perspectiva que este capítulo se insere na discussão. Assim, o objetivo aqui é avaliar a relação entre o registro ignorado da escolaridade do óbito adulto e algumas características sociodemográficas do falecido no Brasil em 2010. Respostas encontradas nesse capítulo serão de grande relevância para definição de parâmetros relacionados a imputação dessas não declarações de escolaridade do óbito que são tratadas no Capítulo 4.

3.1 Metodologia

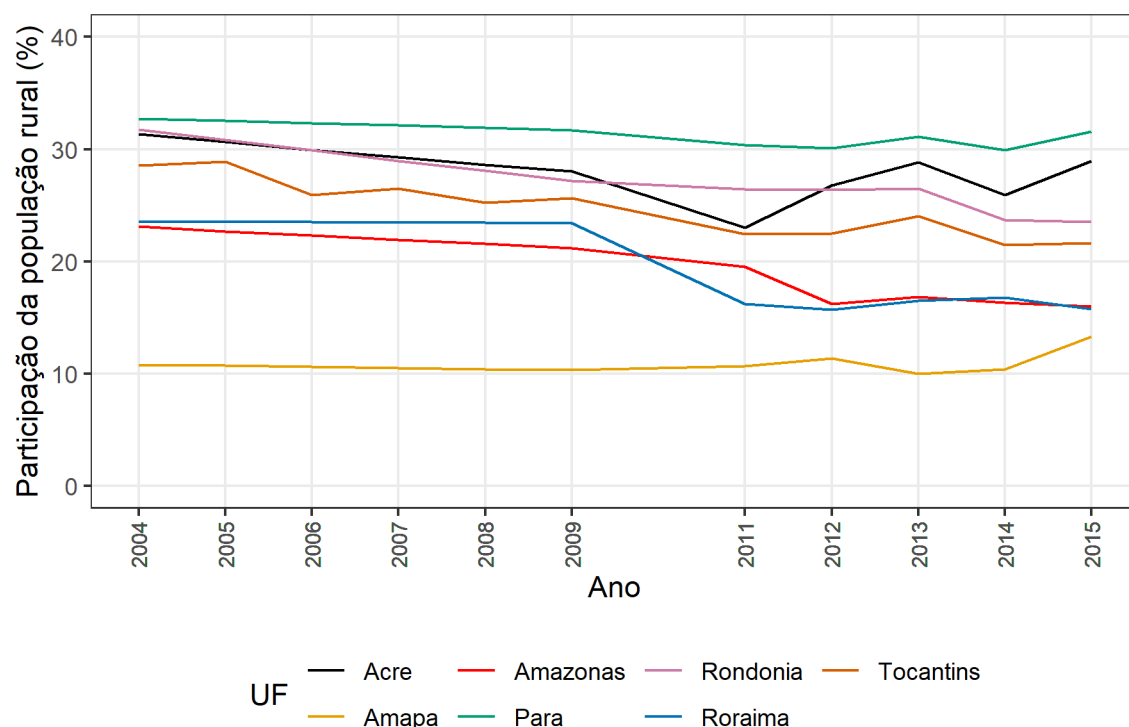
3.1.1 Fonte de dados

O presente estudo utiliza os microdados de óbitos do Sistema de Informação sobre Mortalidade do Ministério da Saúde (MS/DATASUS/SIM) para o Brasil no ano de 2010. Como a variável principal do estudo é a escolaridade, optou-se por utilizar apenas os falecidos com idade de 25 anos a 60 anos (exclusive), pois a partir de 25 anos é razoável aceitar que haja a consolidação do grau de escolaridade de grande parte das pessoas, contemplando inclusive aqueles com defasagem escolar. Embora a chance de escolaridade faltante nos óbitos na idade de 60 anos ou mais seja maior do que entre os adultos de 25 a 59 (SILVA JÚNIOR; FREIRE, 2018), optou-se pela exclusão dos adultos idosos. A justificativa é porque há problemas nas estimativas de mortalidade em idades avançadas, tanto no numerador (óbitos) quanto no denominador (população) (GOMES; TURRA, 2009; QUEIROZ; SAWYER, 2012; QUEIROZ et al., 2017).

A utilização do ano de 2010 se fundamenta pelo fato de ter sido o ano limite em que o Ministério da Saúde passou a recomendar o uso preferencial da nova declaração de óbito que traz mudanças na forma de mensurar a escolaridade, minimizando assim possíveis erros na hora de compatibilizar os dados a serem publicados (MS, 2011; RIBEIRO, 2016). Além disso, essa avaliação fornecerá subsídios para estimar as probabilidades e taxas de morte segundo nível de escolaridade utilizando dados do SIM, uma vez que em 2010 se tem disponível o denominador das taxas, a população recenseada segundo grau de escolaridade. Desta forma realizar este estudo num ano onde se tem dados populacionais censitários por nível de escolaridade facilita o cálculo de taxas de mortalidade segundo essa caracterização.

Os dados mais atuais de óbitos disponíveis no SIM referem-se ao ano de 2016, mas utilizá-los implica na incorporação de mais uma fonte de incerteza no produto final desse trabalho que são as estimativas de mortalidade por escolaridade. Isso porque não se tem disponíveis as quantidades populacionais por escolaridade, sendo necessário projetá-las para o período. Outra alternativa seria utilizar os dados de óbito do SIM do ano de 2014, já que dados populacionais estão disponíveis na PNAD 2014. Embora seja representativa para as UFs, essa pesquisa pode não ser representativa para a área rural da região Norte (PÉREZ; TURRA, 2008). A Figura 1 mostra um comportamento instável da participação da população rural em algumas UFs da região Norte a partir de 2009. No Acre, por exemplo, a participação da população residente com situação de domicílio rural saiu de 28,03% em 2009 para 23,01% em 2011 e voltou a aumentar para 28,83% em 2013. Mais uma queda foi registrada nessa participação em 2014 (25,91%), mas voltou a crescer em 2015 (28,94%).

Figura 1: Participação (%) da população residente nas UF's da região Norte por situação de domicílio rural, 2004 a 2015.

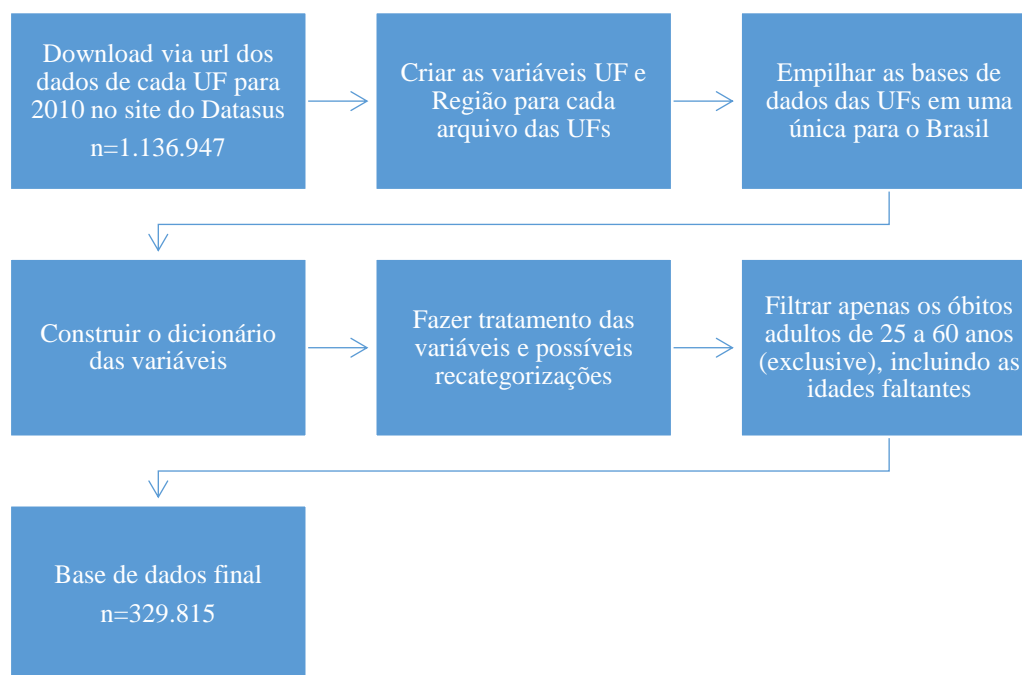


Fonte: IBGE, Pesquisa Nacional por Amostra de Domicílio 2004-2015

Os arquivos dos microdados dos óbitos são disponibilizados de tal forma que cada UF possui um arquivo em formato “.dbc”. A ideia aqui é operacionalizar todos os procedimentos metodológicos realizados nessa dissertação por meio do software R, já que é um software livre e pode facilitar a reprodução dos resultados apresentados. Toda rotina computacional do software R utilizada para gerar a base de dados pode ser obtida no [Repositório da dissertação - scripts em R e resultados](#)². A Figura 2 ilustra o fluxo operacional para obtenção e consolidação dos microdados de óbitos do SIM para 2010.

² Será disponibilizado um repositório de scripts em R e resultados no GitHub do autor (<https://github.com/walterpedro/dissertacao-demografia-mortalidade/blob/master/README.md>).

Figura 2: Fluxo operacional para obtenção e consolidação da base de dados de óbitos registrados pelo Ministério da Saúde/ Datasus / SIM em 2010.



Fonte: Elaboração própria a partir de dados do Ministério da Saúde/Datasus/SIM.

Em relação as variáveis disponíveis, é importante ressaltar que ao todo o SIM disponibiliza em seus microdados 61 variáveis. Muitas delas são preenchidas apenas para óbitos fetais ou para óbitos menores de 1 ano ou para mulheres. Outras são referentes à data, hora e códigos da declaração de óbito. Há ainda aquelas que sequer estão no dicionário³ e que todos os casos são ignorados (ver detalhes na Tabela A1 no apêndice). Essas variáveis não foram consideradas no estudo. Como o interesse é nas características sociodemográficas dos óbitos, as variáveis utilizadas nesta avaliação são a escolaridade, grandes regiões, sexo, idade, raça/cor, estado civil, local de ocorrência e causa de morte. Na Tabela 1, é apresentada a categorização utilizada para cada variável do estudo.

Tabela 1: Variáveis sociodemográficas dos óbitos registrados no SIM em 2010 e categorizações utilizadas.

Variável	Descrição	Categorias dos microdados	Categorias finais
Região	Região de residência do falecido	NA*	1: Norte
			2: Nordeste
Sexo	Sexo do falecido	1: Masculino	3: Sudeste
			4: Sul
			5: Centro Oeste
Sexo	Sexo do falecido	2: Feminino	1: Masculino
		9: Ignorado	2: Feminino
			9: Ignorado

³ Disponível em: ftp://ftp.datasus.gov.br/dissemin/publicos/SIM/CID10/Docs/Estrutura_SIM_para_CD.pdf

Variável	Descrição	Categorias dos microdados	Categorias finais
<i>Continuação (Tabela 1)</i>			
Idade (grupos quinquenais)	Idade do falecido	Variável discreta (em anos)	6: 25 a 29 anos 7: 30 a 34 anos 8: 35 a 39 anos 9: 40 a 44 anos 10: 45 a 49 anos 11: 50 a 54 anos 12: 55 a 59 anos 999: Ignorado (000, 999, em branco)
Cor/raça	Cor ou raça do falecido	000, 999 e em branco: Ignorado 1: Branca 2: Preta 3: Amarela 4: Parda 5: Indígena Em branco	1: Branca 2: Preta 4: Parda 6: Amarela ou indígena 9: Ignorado (Em branco)
Estado civil	Estado civil do falecido	1: Solteiro 2: Casado 3: Viúvo 4: Separado judicialmente 5: 9: Ignorado	1: Solteiro 2: Casado 3: Viúvo 6: Divorciado/Outros (4 e 5) 9: Ignorado
Local de ocorrência	Local de ocorrência do óbito	1: Hospital 2: Outro estabelecimento de saúde 3: Domicílio 4: Via pública 5: Outros 9: Ignorado	1: Hospital 2: Outro estabelecimento de saúde 3: Domicílio 4: Via pública 5: Outros 9: Ignorado
Causa básica de morte (em capítulos)	Causa Básica de morte conforme CID10	Códigos da CID10	2: II. Neoplasias (tumores) 9: IX. Doenças do aparelho circulatório 10: X. Doenças do aparelho respiratório 20: XX. Causas externas 99: Outros
Escolaridade	Escolaridade do falecido (em anos de estudo concluídos)	1: Nenhum 2: 1 a 3 anos 3: 4 a 7 anos 4: 8 a 11 anos 5: 12 anos e mais 9: Ignorado 0: Em branco	1: Nenhum 2: 1 a 3 anos 3: 4 a 7 anos 4: 8 a 11 anos 5: 12 anos e mais 9: Ignorado (9, 0 e em branco)
Escolaridade (3 categorias)	Escolaridade com recategorização em baixa, média e alta escolaridade a partir da variável "Escolaridade"	NA	1, 2, 3 => 1: baixa 4 => 2: média 5 => 3: alta 9 => 9: Ignorado
Escolaridade faltante	Indica se a escolaridade foi missing ou não. Codificada a partir da variável "Escolaridade"	NA	1,2,3,4,5 => 0 : Não 9 => 1: Sim

Fonte: Ministério da Saúde / Datasus

*NA = não se aplica

Conforme ressaltado anteriormente, os microdados são disponibilizados por UF. Com isso, uma variável denominada “UF” foi criada no banco de dados para que fosse possível agregar as respectivas UFs em uma variável denominada “Região”. A variável idade é disponibiliza nos microdados com 3 dígitos, onde foi necessário transformá-los em uma variável discreta (em anos) conforme sugere o dicionário e, posteriormente, em grupos quinquenais. Em relação à variável cor/raça, a categorização original foi mantida, exceto para as categorias “Amarela” e “Indígena”. Elas representam apenas 0,3% e 0,2% dos óbitos adultos, respectivamente, e foram agrupadas. Para a variável estado civil, agregou-se os níveis

“separado judicialmente” (Divorciado) com a categoria “Outros”, as demais foram mantidas. Já para as variáveis sexo e local de ocorrência não se realizou recategorização.

A justificativa para a categorização da variável causa de morte em capítulos, se dá pelo fato dessa variável ser importante no processo de imputação de dados faltantes da escolaridade e pela necessidade de se ter uma categorização com um número de níveis reduzido para viabilizar a aplicação de futuras técnicas de imputação e para otimizar a apresentação dos resultados. Sendo assim, a categorização contempla os capítulos de causa de morte onde a não declaração da escolaridade do óbito foi mais expressiva. Adotou-se um critério de corte de 10%, ou seja, em relação ao total de não declarações de escolaridade, os capítulos de causas de morte com dados ignorados acima desse valor foram tratados individualmente, já categorias abaixo do corte foram agrupadas na categoria "outros capítulos". O valor de 10% foi baseado na classificação da completitude apresentado por Romero e Cunha (2006), onde variáveis com até 10% de registros ignorados são classificadas como excelente ou bom e acima de 10% como regular, ruim ou muito ruim. Outra alternativa para categorização da causa de morte é: causas evitáveis, não-evitáveis e mal definidas (MALTA et al., 2011). Porém, após realizar a categorização dessa maneira, os totais de óbito não corresponderam aos observados nas tabelas do site do Datasus e por isso não foram utilizados.

Baseado na variável escolaridade, criou-se uma variável indicadora chamada “Escolaridade faltante”, em que foi atribuído o valor 1 caso o dado de escolaridade seja ignorado ou em branco e 0 caso contrário. Essa variável é utilizada como a variável resposta aqui. As outras duas formas da variável escolaridade (a original e com 3 categorias) foram incluídas aqui apenas para uma análise descritiva. A categorização da variável Escolaridade (3 categorias) será utilizada para apresentar as estimativas de mortalidade no capítulo 6, já que a literatura aponta que foram observados resultados mais consistentes com um menor número de categorias e pela praticidade de representá-las nos resultados (RIBEIRO, 2016). Cabe ressaltar aqui que as pessoas em nível superior estarão agrupadas numa categoria única, seja para quem está cursando um curso de nível superior, seja para aqueles que já concluíram essa etapa. Por outro lado, a literatura também chama atenção para o fato de utilizar a última categoria com intervalo aberto (12 anos de estudo ou mais), pois indivíduos que obtiveram diplomas de mestrado e doutorado podem ter taxas de mortalidade ainda menores do que indivíduos cujo nível educacional mais alto foi a graduação (HUMMER; HERNANDEZ, 2013).

É importante ressaltar que, além dos dois problemas de qualidade apresentados aqui nesse trabalho (incompletitude e de cobertura), os microdados de óbitos do SIM carece de

padronização e de um maior controle da tabulação de algumas variáveis. De maneira mais detalhada, a Tabela A2 do apêndice apresenta a descrição dessas variáveis, bem como a categorização original proposta no dicionário, a codificação nos microdados, as categorias publicadas nas tabelas consultas no site⁴ do Datasus e a categorização utilizada. Para algumas variáveis, observa-se uma discordância entre as categorias propostas no dicionário, as que se observa nos microdados e as que o Datasus disponibiliza nas tabelas consultadas no site. Na variável idade, o dicionário estabelece o código 000 para dados ignorados, porém nos microdados aparecem três códigos para os registros ignorados (000, 999 e caselas em branco).

A variável cor/raça também tem desconformidade. O dicionário não estabelece um código para dados ignorados, porém nos microdados esses foram deixados em branco. Outro problema foi encontrado na variável estado civil. Nesse atributo a codificação “5” consta nos microdados, mas não aparece no dicionário. Ao consultar as tabelas do site do Datasus, esse código foi considerado como a categoria “outro”.

Situação semelhante se encontra a principal variável do estudo: escolaridade. Nos microdados aparece o código “0”, mas não está contemplado no dicionário. Também houve registro de caselas em branco para a escolaridade do falecido. As tabelas oficiais do site consideraram o código “9”, “0” e “em branco” como dado faltante. As demais variáveis utilizadas aqui não apresentaram problemas dessa natureza.

Essas inconsistências representam uma possível fonte de erro, principalmente para a variável escolaridade. O código “0” foi considerado como ignorado, mas não tem como saber se esses zeros não eram para ser contabilizados na categoria “nenhuma” escolaridade, como sendo 0 (zero) anos de estudo. Isso pode subestimar as quantidades de óbitos nos menos escolarizados e, conseqüentemente, subestimar as estimativas de mortalidade dos menos instruídos.

Também é importante alertar para possíveis problemas com a informação prestada sobre a escolaridade do falecido em declarações de óbitos (DO). Apesar de ser obrigatório apenas a partir de 2011, o novo modelo de declaração de óbito pode estar sendo utilizado simultaneamente em UFs diferentes desde 2009. Isso pode ser uma fonte de erro, já que na DO nova o quesito escolaridade sofreu modificação. O Ministério da Saúde compatibiliza as duas antes de disponibilizar os microdados (MS, 2011; RIBEIRO, 2016). Outra fonte de erro é o fato de as informações da DO serem dadas por terceiros que muitas vezes podem não ter

⁴ Site do Datasus para consulta das tabelas: <http://tabnet.datasus.gov.br/cgi/defthtm.exe?sim/cnv/obt10uf.def>

conhecimento suficiente sobre a vítima (PÉREZ, 2010; GUEDES et al., 2011; RIBEIRO, 2016). Estudos americanos mostram que terceiros têm uma preferência por relatar escolaridades maiores para os falecidos em atestados de óbito, o que pode subestimar as menores escolaridades (HUMMER; HERNANDEZ, 2013). Porém, esse problema é mais evidente em idosos, o que não afeta muito o recorte aqui estudado (SORLIE; JOHNSON, 1996).

Outra opção de fontes de dados de óbitos para estudos sobre diferenciais educacionais na mortalidade é o Censo de 2010 (RIBEIRO, 2016), mas apenas sexo e idade das pessoas que morreram foram inquiridas. Assim, trabalhos que utilizaram essa fonte estimaram de maneira indireta esses gradientes educacionais. Apesar das limitações apresentadas, a base de dados do SIM é a única no Brasil que permite estudar esses diferenciais a partir do registro individual da escolaridade da pessoa falecida.

3.1.2 Método

Vale salientar que o objetivo dos métodos empregados neste capítulo não é investigar diferenciais de mortalidade, mas sim de caracterizar a distribuição dos eventos de óbito visando os dados ignorados, ou seja, a completitude. Sabe-se também que os óbitos possuem sub-registros diferenciado por cada variável apresentada, mas a análise e o tratamento desses fogem do escopo aqui e são tratados no Capítulo 5. No tocante à análise dos dados, empregou-se análise descritiva univariada e bivariada por meio de frequências absolutas (n) e relativas (%). No primeiro momento, analisou-se, de maneira univariada, como os dados faltantes na variável escolaridade do óbito adulto se comportaram por região ao longo de todo período de disponibilidade dos dados do SIM à época da pesquisa – 1979 a 2016. Com foco no ano de 2010, uma análise descritiva univariada das variáveis explicativas, além da escolaridade, é realizada. De maneira bivariada, as variáveis explicativas são analisadas segundo a variável resposta escolaridade faltante.

Para identificar a relação entre a variável resposta Escolaridade ignorada (Sim/Não) e as variáveis explicativas (sociodemográficas) utilizou-se como medida de associação a Razão de Chances (RC). Aqui é importante ressaltar que, como os dados de óbito do SIM são de universo e não de amostra, apenas análises descritivas foram realizadas, não cabendo, portanto, análises inferenciais como modelos de Regressão Logística ou de Poisson, por exemplo.

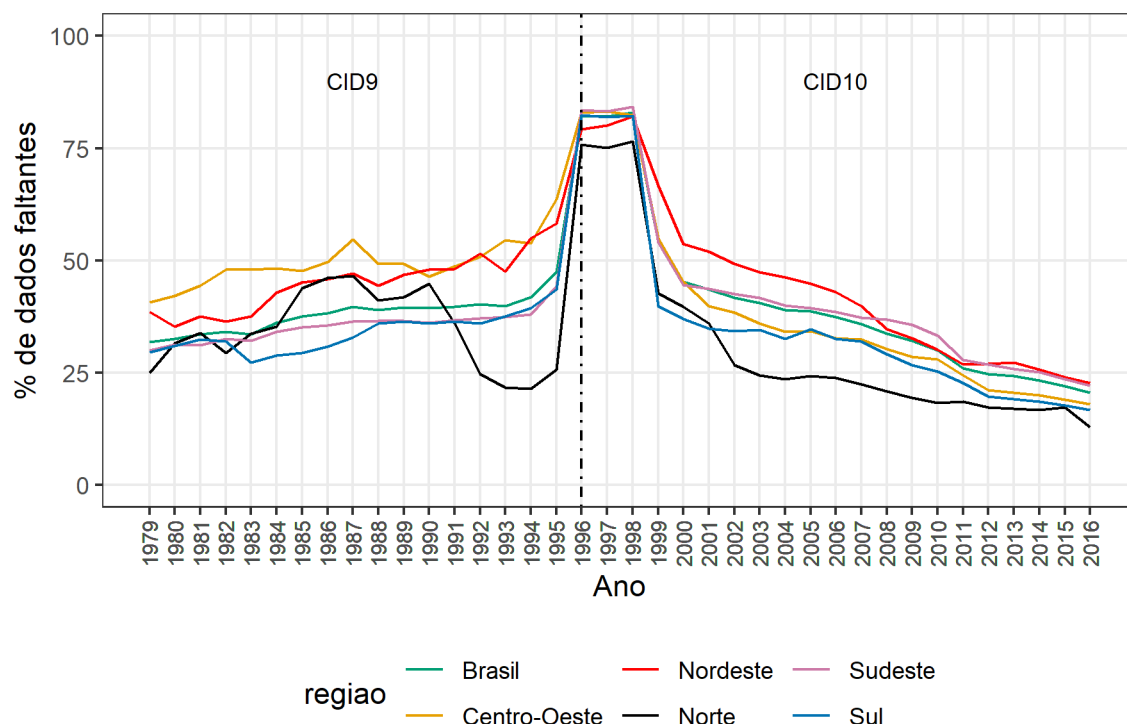
Todas as análises foram realizadas por meio do software R e as rotinas estão disponíveis no [Repositório da dissertação - scripts em R e resultados](#).

3.2 Resultados

Ao realizar uma análise preliminar do histórico da qualidade dos dados na perspectiva da completitude do quesito “Escolaridade” do óbito adulto (incluindo os casos com idade faltante) registrados no SIM, verificou-se que os percentuais de dados faltantes nessa variável de 1979 a 2016 (exceto 1996 a 1998) para o Brasil variaram de 21% a 54%. Nos anos de 1996 a 1998 ocorreu uma negligência no preenchimento das declarações de óbito, chegando a um percentual médio de dado faltante de 82% no período. Fica evidenciado também que esse descaso ocorreu em todas as regiões brasileiras no mesmo período. Não foram encontrados os motivos, mas provavelmente isso ocorreu em função do despreparo dos profissionais da saúde frente às mudanças ocorridas da CID-9 para CID-10 de 1995 para 1996, ressalvado mera coincidência.

No que se refere à evolução da completitude do registro, após 1998 observa-se uma tendência de melhoria da qualidade desse quesito em todas as regiões brasileiras, saindo de patamares de 70 a 80% de dados faltantes em 1998 para 10% a 20% em 2016. Surpreendentemente, um destaque positivo é a região Norte, onde desde 2002 é a que apresentou menor percentual de registros com a escolaridade ignorada. Em 2010, que é o foco desse estudo, foram registrados 329.815 óbitos adultos (incluindo os óbitos com idades faltantes) no Brasil. Desse total, o percentual de ausência de registro da escolaridade do óbito em idade adulta foi de 29,96% (98.820 casos). As regiões Centro-Oeste, Sul e Norte apresentaram percentuais de registros ausentes abaixo do nível do Brasil para o ano de 2010, com 27,11%, 24,79% e 17,42%, respectivamente, enquanto o Nordeste (30,19%) e Sudeste (33,29%) ficaram acima (Figura 3).

Figura 3: Percentuais de dados faltantes na variável escolaridade do óbito de 25 a 59 anos incluindo os casos com a idade faltante, Brasil, grandes regiões, 1979 a 2016.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 1979 a 2016 (consulta do site).

É importante lembrar aqui que em todas as análises desse capítulo nas quais utilizaram dados oriundos dos microdados, os dados faltantes das variáveis sociodemográficas foram considerados como uma categoria a mais. Sendo assim, o total de casos de óbitos foi de 329.815. A Tabela 2 permite avaliar a distribuição dos óbitos univariados das variáveis de interesse no trabalho.

Em relação à distribuição dos óbitos por região, verifica-se que não houve casos ignorados, pois os microdados foram extraídos por unidade da federação, o que permitiu a criação da variável região sem perda de registro. As maiores participações nos óbitos são da região Sudeste (46,3%), seguida da Nordeste (24,9%). A causa básica de morte também não apresentou dados ignorados. As causas de morte mais frequentes foram as do capítulo XX que trata das causas externas (por exemplo, acidente de trânsito, homicídio) com 23,9% dos casos. Os capítulos IX das doenças do aparelho circulatório e II das Neoplasias (tumores) também foram bem representativas, com percentuais de 21,3% e 17,3%, respectivamente.

Todas as outras variáveis apresentaram dados faltantes: na variável sexo, com 0,1% de dados ignorados; idade, com 1,2% de registros sem declaração; local de ocorrência do óbito, com 0,4% de registros sem essa informação; cor/raça com 6,3% de não informação e; estado

civil com 8,4% de casos sem essa informação. Já a variável escolaridade apresentou um percentual de 30,0% de declarações de óbitos sem informação. No tocante à distribuição das outras categorias, os óbitos são mais frequentes no sexo masculino (67,2%). O quantitativo de óbitos é maior a medida em que aumenta a idade, acometem mais os brancos (44,0%) e pardos (39,8%), os solteiros (43,4%) e casados (36,6%), e a maioria (63,4%) ocorrem em hospitais. Em relação à distribuição dos óbitos segundo a escolaridade, observa-se uma grande concentração de óbitos nas menores faixas de anos de estudo, onde quase a metade dos falecidos possui até 7 anos de estudo. Apenas 6,4% dos que morreram em 2010 possuíam 12 anos ou mais de estudo.

Tabela 2: Distribuição do número de óbitos adultos no Brasil em 2010 segundo variáveis demográficas e escolaridade.

Variáveis	Categorias	n	%
Região	Norte	21619	6,6
	Nordeste	82048	24,9
	Sudeste	152661	46,3
	Sul	49647	15,1
	Centro Oeste	23840	7,2
Sexo	Masculino	221753	67,2
	Feminino	107858	32,7
	Ignorado	204	0,1
Idade	25 a 29 anos	28769	8,7
	30 a 34 anos	29916	9,1
	35 a 39 anos	32089	9,7
	40 a 44 anos	40549	12,3
	45 a 49 anos	52849	16,0
	50 a 54 anos	65460	19,8
	55 a 59 anos	76216	23,1
	Ignorado	3967	1,2
Cor/raça	Branca	144998	44,0
	Preta	31000	9,4
	Parda	131370	39,8
	Amarela ou indígena	1590	0,5
	Ignorado	20857	6,3
Estado civil	Solteiro	143070	43,4
	Casado	120659	36,6
	Viúvo	12856	3,9
	Divorciado/Outros	25422	7,7
	Ignorado	27808	8,4
Local de ocorrência	Hospital	209194	63,4
	Outro estabelecimento de saúde	11118	3,4
	Domicílio	55177	16,7
	Via pública	35534	10,8
	Outros	17412	5,3
	Ignorado	1380	0,4
Causa básica de morte (em capítulos)	II. Neoplasias (tumores)	56900	17,3
	IX. Doenças do aparelho circulatório	70246	21,3
	X. Doenças do aparelho respiratório	18898	5,7
	XX. Causas externas de morbidade e mortalidade	78935	23,9
	Outros	104836	31,8

Variáveis	Categorias	n	%
<i>Continuação (Tabela 2)</i>			
Escolaridade	Nenhuma	24912	7,6
	1 a 3 anos	56381	17,1
	4 a 7 anos	82220	24,9
	8 a 11 anos	46294	14,0
	12 anos e mais	21188	6,4
	Ignorado	98820	30,0
Escolaridade (com 3 categoria)	Baixa	163513	49,6
	Média	46294	14,0
	Alta	21188	6,4
	Ignorado	98820	30,0
Escolaridade faltante	Não	230995	70,0
	Sim	98820	30,0
Total (N)		329815	100,0

Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010.

Após um panorama geral da distribuição univariada dos óbitos por cada variável considerada no estudo, a Tabela 3 fornece as frequências absolutas (n), relativas (%) e razões de chance (RC) da variável resposta escolaridade faltante por cada variável explicativa. Ao analisar por região, verifica-se que o maior percentual de óbitos com escolaridade faltante foi a região sudeste (33,3%) enquanto que o menor foi a região norte (18,3%). Com relação ao sexo do falecido, quase não se observa diferença nos percentuais, onde óbitos do sexo masculino e do sexo feminino apresentaram 30,7% e 28,3% de dados faltantes de escolaridade. Dos casos onde o sexo é *missing*, 94,6% também não tiveram a escolaridade registrada. Em relação à idade, observa-se uma uniformidade de dados faltantes em torno de 29% em cada faixa etária. Dos óbitos com idade não declarada, 97,5% não tiveram a escolaridade preenchida.

Ainda na Tabela 3, quando se analisa cor ou raça, o percentual de escolaridade faltante entre os pretos foi de 28,7%, ao passo que o menor percentual foi entre os amarelos ou indígenas, com 25,0%. Em 66,2% dos casos em que a variável cor ou raça foi ignorada, a informação sobre a escolaridade foi perdida. Os solteiros possuem o menor percentual de escolaridade ignorada (24,2%), ao passo que os viúvos são os que mais apresentaram o problema (26,5%). Entre os casos com o estado civil faltante, 83,7% estavam sem o nível de instrução. Já em relação ao local de ocorrência do óbito, dos óbitos que ocorreram em via pública 35,5% apresentaram dado faltante na escolaridade, dos que faleceram no hospital esse percentual foi de 29,0%, e, com menor percentual, 21,7% dos que morreram em outro estabelecimento de saúde não tiveram sua escolaridade registrada. Ainda sobre o local de ocorrência, entre aquelas declarações de óbito onde essa variável foi ignorada, 61,3% ficaram sem registro da escolaridade. Ao desagregar por causa básica de morte, verifica-se que os percentuais de escolaridade faltante foram maiores no capítulo XX – Causas externas com 33,3%.

Tabela 3: Distribuição e Razões de Chance (RC) da variável Escolaridade ignorada dos óbitos adultos segundo variáveis sociodemográficas, Brasil, 2010.

Variáveis	Categorias	Escolaridade faltante				RC*
		Não		Sim		
		<i>n</i>	%	<i>n</i>	%	
Região (Norte)	Norte	17655	81,7	3964	18,3	1,000
	Nordeste	57275	69,8	24773	30,2	1,926
	Sudeste	101833	66,7	50828	33,3	2,223
	Sul	37084	74,7	12563	25,3	1,509
	Centro Oeste	17148	71,9	6692	28,1	1,738
Sexo (Feminino)	Masculino	153605	69,3	68148	30,7	1,126
	Feminino	77379	71,7	30479	28,3	1,000
	Ignorado	11	5,4	193	94,6	44,544
Idade (25 a 29 anos)	25 a 29 anos	20608	71,6	8161	28,4	1,000
	30 a 34 anos	21062	70,4	8854	29,6	1,062
	35 a 39 anos	22707	70,8	9382	29,2	1,043
	40 a 44 anos	28569	70,5	11980	29,5	1,059
	45 a 49 anos	37411	70,8	15438	29,2	1,042
	50 a 54 anos	46392	70,9	19068	29,1	1,038
	55 a 59 anos	54147	71,0	22069	29,0	1,029
	Ignorado	99	2,5	3868	97,5	98,661
Cor ou Raça (Branca)	Branca	103705	71,5	41293	28,5	1,000
	Preta	22116	71,3	8884	28,7	1,009
	Parda	96942	73,8	34428	26,2	0,892
	Amarela ou indígena	1192	75,0	398	25,0	0,839
	Ignorado	7040	33,8	13817	66,2	4,929
Estado civil (Solteiro)	Solteiro	108501	75,8	34569	24,2	1,000
	Casado	89665	74,3	30994	25,7	1,085
	Viúvo	9452	73,5	3404	26,5	1,130
	Divorciado/Outros	18854	74,2	6568	25,8	1,093
	Ignorado	4523	16,3	23285	83,7	16,158
Local de ocorrência (Domicílio)	Hospital	148441	71,0	60753	29,0	1,050
	Outro estabelecimento de saúde	8108	72,9	3010	27,1	0,952
	Domicílio	39698	71,9	15479	28,1	1,000
	Via pública	22903	64,5	12631	35,5	1,414
	Outros	11311	65,0	6101	35,0	1,383
	Ignorado	534	38,7	846	61,3	4,063
Causa básica de morte (II. Neoplasias (tumores))	II. Neoplasias (tumores)	41954	73,7	14946	26,3	1,000
	IX. Doenças do aparelho circulatório	50320	71,6	19926	28,4	1,112
	X. Doenças do aparelho respiratório	13139	69,5	5759	30,5	1,230
	XX. Causas externas	52681	66,7	26254	33,3	1,399
	Outros	72901	69,5	31935	30,5	1,230

Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010.

*A variável resposta é Escolaridade ignorada (Sim/Não) onde o desfecho é o Sim para o cálculo da RC.

Ainda no tocante à Tabela 3, com relação à variável região, constata-se que todas as outras regiões possuem uma maior chance de se observar dados faltantes de escolaridade do óbito comparado à região Norte. Destacam-se os óbitos registrados na região Sudeste que têm uma chance 2,223 vezes de apresentar dado faltante na variável escolaridade do que os registrados na região Norte. Com relação ao sexo, a chance de escolaridade faltante nos óbitos do sexo masculino é 1,126 vezes comparada aos do sexo feminino. Para idade, todas as faixas etárias apresentaram uma maior chance de apresentarem dados de escolaridade ignorado do que os falecidos mais jovens de 25 a 29 anos, com razões de chance variando na ordem de 1,029 a 1,062 vezes.

Já em relação às diferenças entre raça ou cor do falecido (Tabela 3), com exceção da cor ou raça preta (que praticamente não houve diferença entre as chances), todas as categorias apresentaram menores chances de apresentar dados faltantes na escolaridade comparadas aos brancos, sendo a maior delas entre amarela ou indígena e brancos, com chance 1,192 vezes menor ($RC = 1/0,839$) nos óbitos declarados como amarela ou indígena. Analisando sob o ponto de vista do estado civil do falecido, observa-se que as chances desse desfecho ocorrer nos casados, viúvos e divorciados/outros são 1,085, 1,130 e 1,093 vezes maiores comparados aos falecidos que eram solteiros. As chances também diferem quando se leva em consideração o local de ocorrência, com destaque para os que ocorreram em via pública. Esses têm uma chance 1,414 vezes maior de terem a escolaridade ignorada no momento do cadastro comparado àqueles óbitos ocorridos no domicílio do falecido, já quando o óbito ocorre no hospital, essa chance é 1,050 vezes maior. Quando se leva em consideração as causas básicas de morte, os óbitos por causas externas são os que apresentam maior chance (1,399 vezes) de ocorrência de ausência no registro da escolaridade comparados aos óbitos por Neoplasias (tumores).

É fundamental chamar atenção para o comportamento das chances de se observar a escolaridade faltante quando também não se observam as variáveis explicativas. Elas são extremamente maiores comparadas às categorias de referência de cada quesito. Para a variável sexo esse valor foi 44,564 vezes. Para idade (98,661), cor ou raça (4,929), estado civil (16,158) e local de ocorrência (4,063).

3.3 Discussão

Por meio das análises realizadas, percebeu-se que as chances de a escolaridade do óbito ser ignorada no ato do preenchimento da declaração de óbito diferem segundo características sociodemográficas. No que tange à variável região, verificou-se que a chance de apresentar

dados faltante é menor na região norte comparado com as demais regiões. Esse resultado é contrário a todos os estudos que avaliaram a qualidade dos dados sob o aspecto do grau de cobertura dos óbitos por região. Queiroz et al. (2017) verificaram que, em 2010, todos os estados das regiões Sul e Sudeste apresentaram cobertura de 100% dos registros de óbitos do SIM (pelo método GGB), ao passo que em alguns da região Norte detectou-se sub-registro. Com isso, depreende-se que uma melhor cobertura não significa melhor qualidade dos dados sob a dimensão completitude.

Verificou-se ainda que os homens, em idades mais avançadas, brancos, viúvos, que morreram em via pública e por causas externas têm uma maior chance de apresentar registro de escolaridade ignorado na declaração de óbito comparado à categoria de referência de cada uma das variáveis (conforme já apresentado na Tabela 3). As chances também foram extremamente maiores em todos os casos quando as variáveis explicativas foram ignoradas. Essa concomitância de dados ignorados em mais de uma variável do estudo pode representar uma limitação para o uso de métodos que visem imputar a variável escolaridade baseado na distribuição de variáveis auxiliares. Alguns deles exigem que essas variáveis sejam completas, ou seja, sem dados faltantes.

Romero e Cunha (2006) chamam atenção para um fator que, de maneira geral, possa ter relação com essas diferenças. Eles alegam que maiores incompletitudes observadas podem ser influenciadas pelo grau de cobertura diferenciados nessas variáveis, com o argumento de que em UFs com menor grau de cobertura (como é o caso da região Norte) tendem a ter, possivelmente, maiores proporções de declarações de óbito preenchida por médicos e agentes administrativos em melhores condições.

Os resultados desse capítulo podem ser considerados como uma análise prévia para tentar compreender o comportamento dos dados faltantes de escolaridade do óbito, objetivando a realização de estimativas de mortalidade adulta segundo o nível de escolaridade aplicando métodos de imputação para corrigir essa incompletitude, contribuindo, assim, para a compreensão das desigualdades educacionais presentes nas probabilidades de morte e no tempo médio em que a população espera viver.

Vale chamar atenção, também, para alguns estudos que fizeram uso de técnicas inferenciais quando analisam a completitude de variáveis do SIM. Por exemplo, Felix et al. (2012) e Macente e Zandonade (2010) utilizaram modelos de regressão para avaliar tendência na completitude de variáveis sociodemográficas. Porém, entende-se aqui que técnicas

estatísticas inferenciais como modelos de regressão linear, logística ou de Poisson não se aplicam nesse tipo de estudo, pois os dados de óbito do SIM devem ser tratados como dados de universo e não de amostra. A não ser que seja adotada uma abordagem de superpopulações, que não foi o caso. Por esse motivo, aqui neste trabalho não se utilizou, por exemplo, modelos de regressão logística para estimar as razões de chances e seus intervalos de confiança.

No tocante aos desdobramentos da discussão sobre a qualidade das informações dos sistemas de saúde, uma abordagem que é bastante comum na literatura para a melhoria dessa qualidade e que merece ser destacada é a importância de um esforço maior na capacitação dos profissionais de saúde responsáveis pelo preenchimento das declarações de óbito, incluindo alterações no currículo dos cursos de medicina, visando o ensino do preenchimento e que mostre a relevância do preenchimento por completo e correto (STUQUE et al., 2003; ROMERO; CUNHA, 2006; COSTA; FRIAS, 2011).

Por fim, espera-se que, além da importância para aplicação de métodos de imputação aplicados nesse trabalho, esse capítulo venha a contribuir para uma discussão sobre o que pode estar determinando esses dados ignorados na escolaridade do óbito.

4 IMPUTAÇÃO DE DADOS FALTANTES NA ESCOLARIDADE DO ÓBITO NO SIM

Este capítulo tem como objetivo apresentar uma breve revisão da utilização de métodos de imputação em estudos sobre diferenciais de mortalidade no Brasil, bem como definir algumas premissas conceituais e metodológicas para aplicação de um método de imputação para corrigir os dados faltantes da variável escolaridade do óbito registrados no SIM em 2010. Além disso, os resultados da imputação serão apresentados e comparados com os dados sem imputação. Adicionalmente, pretende-se realizar simulações para validação do método utilizado.

Conforme demonstrado no Capítulo 3, a variável escolaridade dos óbitos adultos registrados no SIM em 2010 possui uma alta porcentagem de dados faltantes e que essas ausências de registro estão associadas a variáveis sociodemográficas do próprio falecido. Esses valores faltantes impedem que estimativas de mortalidade por escolaridade sejam realizadas com métodos padrões sem que antes essa problemática seja resolvida. Caso essa questão não seja levada em consideração, estimativas enviesadas podem ser produzidas e conclusões equivocadas retiradas. Isso porque a distribuição dos dados faltantes na variável pode ser bastante diferente da distribuição dos que tiveram o registro preenchido (RUBIN, 1987). Por exemplo, se os registros faltantes de escolaridade não forem imputados, há uma chance de as estimativas de mortalidade para os falecidos menos escolarizados estarem subestimadas, supondo que uma pior qualidade dos dados esteja relacionada com piores condições socioeconômicas. Além disso, as chances de ausência são maiores para determinados grupos de idade, sexo e regiões, onde poderia resultar em quantidades de óbito subestimadas por essas variáveis.

Em países desenvolvidos, há poucos trabalhos que fizeram o uso de técnicas de imputação para corrigir os dados antes de realizar estimativas dos diferenciais de mortalidade por escolaridade. Isso porque a maioria das pesquisas são realizadas por meio de dados longitudinais prospectivos, em que o indivíduo é inquirido sobre sua escolaridade ainda vivo e são acompanhados até o evento morte. Os dados básicos sobre escolaridade utilizado nessas pesquisas já são imputados antes da divulgação. Embora sejam de excelente qualidade sob a dimensão completitude, a variável escolaridade da *Current Population Survey (CPS)*, por exemplo, apresentou 2,6% de dados faltantes em 2003 e 2,8% em 2009. Esses dados ignorados foram tratados pelo método de imputação *Hot Deck* (ROSTRON et al., 2010).

A necessidade de imputação de informações faltantes é maior em países de pior qualidade dos dados de óbitos, geralmente, países em desenvolvimento. No estudo de Sandoval e Turra (2015) para o Chile, observou-se que a variável escolaridade possuía 0,88% de *missing* no período de 1991-1993 e de 0,11% em 2001-2003, mas por serem percentuais baixos os autores não fizeram imputação e apenas excluíram esses casos da base de dados. Na Argentina, em torno de 40% dos óbitos apresentaram escolaridade faltante em 2010. Para contornar o problema, o autor utilizou a técnica de Imputação Múltipla (IM) antes de prosseguir no estudo sobre o diferenciais de mortalidade. Porém, não entrou em detalhes sobre o método utilizado para gerar as múltiplas bases de dados completos (MANZELLI, 2014).

Se estudos sobre diferenciais de mortalidade por nível de escolaridade são escassos no Brasil, aqueles que utilizaram técnicas de imputação para estudar esta componente demográfica são ainda mais. Essas técnicas são utilizadas com mais frequência na área da saúde, mas em pequenos grupos de indivíduos (NUNES, 2007; NUNES et al., 2009; LEVY et al., 2010; NUNES et al., 2010; CAMARGOS et al., 2011; OLIVEIRA et al., 2018). No contexto demográfico e estudos sobre diferenciais de mortalidade por escolaridade, apenas quatro trabalhos fizeram uso de técnicas de imputação de dados faltantes para corrigir a incompletude na variável utilizada para representar a escolaridade do indivíduo morto (PÉREZ; TURRA, 2008; PÉREZ, 2010; RIBEIRO, 2016; TURRA et al., 2018). Pérez e Turra (2008) e Perez (2010) foram pioneiros na utilização dessas técnicas de imputação em estudos de diferenciais educacionais na mortalidade no Brasil. Além de realizarem uma revisão conceitual e dos métodos de imputação, os autores realizaram correção para cerca de 16,6% de registro da escolaridade da mãe faltante na PNAD 1996 e 3,4% na Pesquisa sobre Padrões de Vida (PPV) que foi utilizada como *proxy* para escolaridade do falecido.

Outro estudo recente que utilizou esses métodos antes de estimar mortalidade adulta por escolaridade é Ribeiro (2016). Semelhante ao trabalho Pérez (2010), uma ampla revisão dos conceitos e dos métodos de imputação foram realizadas de maneira louvável. A autora utilizou esses métodos de uma maneira não muito usual para imputar todas as observações de escolaridade dos indivíduos falecidos pesquisados no Censo 2010 para o estado de São Paulo, baseando-se na escolaridade dos sobreviventes. Isso porque apenas sexo e idade do falecido foram inquiridos. Agregando os falecidos e os sobreviventes, os percentuais de perda foram pequenos, variando entre 0,064% e 1,063%. Três métodos distintos foram comparados. O primeiro é o método *Hot Deck* com seleção via *ABB (Approximate Bayesian Bootstrap)*. Outro método testado foi a Imputação Múltipla com regressão logística ordinal. Por fim, o Modelo

Multivariado Normal (MVN) através do algoritmo EM e Monte Carlo via Cadeia de Markov (MCMC) foram utilizados para imputação. Após várias simulações para verificar a consistência das imputações e baseado em comparações com os dados do SIM como referência, o modelo MVN foi o que produziu distribuições da escolaridade do óbito do Censo mais consistente e adequado (RIBEIRO, 2016).

Conforme destacado, o trabalho de Ribeiro (2016) utilizou dados de óbito por escolaridade do município de São Paulo registrados no SIM em um período entre 2012 e 2013 como padrão ouro para comparação com os resultados obtidos para o Censo. Apesar de apresentar um grau de cobertura de 100% (QUEIROZ, 2012), os dados do SIM para São Paulo ainda assim precisaram de correção de dados faltantes na variável escolaridade, por apresentarem 7,7% de dados faltantes. Nesse caso, foi empregado imputação simples do método *Hot Deck*. Esse é o primeiro registro da utilização dos dados do SIM para estudos de diferenciais educacionais na mortalidade.

O trabalho mais recente nesse contexto de imputação e diferenciais de mortalidade por escolaridade é o de Turra et al. (2018). Eles utilizaram o método *Hotdeck* com seleção estocástica via aproximação bayesiana por amostragem *bootstrap* (ABB) para completar a base de dados de óbitos do SIM do ano de 2014. O foco do estudo foi de estimar os diferenciais de mortalidade por escolaridade segundo sexo e idade apenas nas capitais em que as ausências na escolaridade não ultrapassassem 20%. O que chamou atenção na metodologia empregada foi que não se utilizou a imputação múltipla no processo. Isso pode resultar em conclusões equivocadas acerca dos gradientes encontrados, pois não se levou em consideração a variabilidade devida ao processo aleatório de imputação.

Diante disso, é necessário avançar na utilização da principal fonte de mortalidade do país, que é o SIM, para estimar esses diferenciais em outros recortes mais amplos, como para o Brasil e grandes regiões, segundo sexo e idade.

4.1 Metodologia

4.1.1 Fonte de dados

Os dados utilizados nesse capítulo são os mesmos apresentados no Capítulo 3, ou seja, óbitos adultos (25 a 59 anos) segundo escolaridade, região, sexo, idade, cor/raça, estado civil, local de ocorrência e causa básica de morte (CID10). Embora todas as variáveis tenham apresentado chances diferenciadas de registro faltante na variável escolaridade entre suas categorias, faz-se necessário uma avaliação mais aprofundada de quais variáveis considerar no

processo de imputação. Essa seleção leva em consideração as exigências do método escolhido e será melhor discutida no tópico 4.1.2.3. Antes de apresentar os critérios utilizados para a escolha dessas variáveis (além das razões de chance do Capítulo 3), faz-se necessário compreender os mecanismos geradores desses dados faltantes, bem como os aspectos teóricos acerca dos métodos disponíveis.

4.1.2 Método

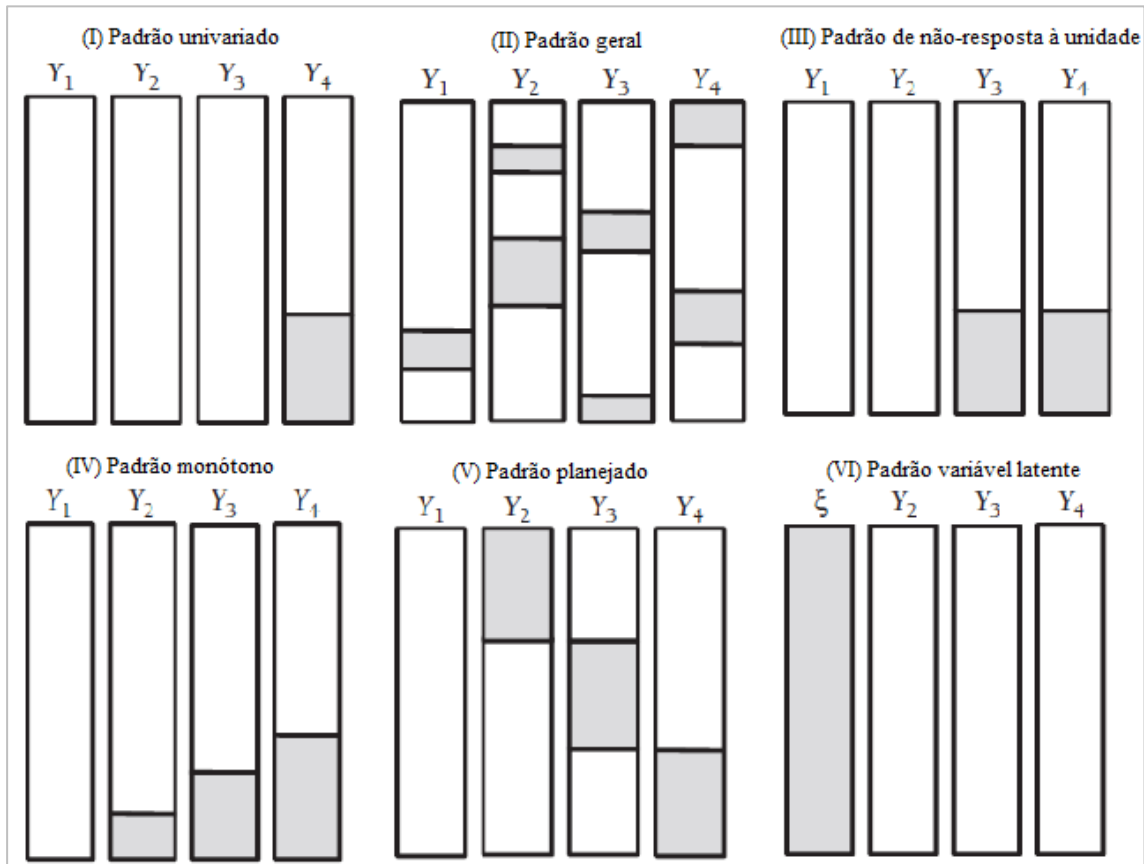
4.1.2.1 Padrões e mecanismos geradores de dados faltantes

Em estudos que envolvem processo de imputação da ausência de resposta a um item é de suma importância diferenciar os conceitos de padrão de dados faltantes e mecanismos que os geram. Padrão de dados ausentes refere-se à configuração dos valores observados e ausentes em um conjunto de dados, enquanto os mecanismos de dados ausentes descrevem possíveis relações (não causal) entre variáveis observadas e a probabilidade de se registrar dados ausentes (ENDERS, 2010).

Em relação ao padrão de dados faltantes, seis tipos são apresentados por Enders (2010) e são mostrados pela Figura 4. O tipo I é chamado de padrão univariado. Nesse caso, dados faltantes são observados em apenas uma variável do estudo. No tipo II, denominado padrão geral, dados faltantes estão presentes em todas ou na maioria das variáveis do conjunto de dados e aparentemente (pode haver relação entre uma variável e a ausência de outra) estão dispostos de maneira aleatória. O autor argumenta ser esse o padrão mais comum na prática. O padrão III está relacionado a casos em que os respondentes se recusaram a responder todos os itens da pesquisa e geralmente ocorrem em pesquisas do tipo *survey*. Em pesquisas longitudinais, pode ocorrer a desistência da participação de alguns entrevistados e pode gerar dados ausentes com a configuração do padrão IV. Já o V é o padrão planejado. Esse tipo de configuração é produzido intencionalmente e são úteis para coletar um grande número de variáveis do questionário quando se deseja reduzir o custo por unidade respondente. Por último, o padrão VI é observado exclusivamente em estudos que utilizam modelos de equação estruturais (análise fatorial confirmatória), onde as variáveis latentes são faltantes para toda a amostra (ENDERS, 2010).

Trazendo para a realidade da pesquisa, no Capítulo 3 ficou claro que os dados se encaixam no padrão geral, já que dados faltantes foram observados em várias variáveis na base de dados do SIM. Porém, como o interesse é na imputação apenas da variável escolaridade, será considerado que a configuração dos dados é dada pelo padrão univariado.

Figura 4: Tipos de padrão de dados faltantes.



Fonte: Adaptado de Enders (2010)

As definições, conceitos e tipos de mecanismos geradores de dados faltantes que são amplamente utilizados na literatura foram desenvolvidos por Rubin (1976) e Little e Rubin (1987). Compreender esses mecanismos de dados faltantes é fundamental, pois as propriedades dos métodos de dados faltantes dependem muito da natureza das dependências desses mecanismos. Além disso, é necessário supor que os dados ausentes foram gerados por um dos mecanismos para definição do método de imputação utilizado (LITTLE; RUBIN, 2002; ENDERS, 2010). Toda teoria e conceitos sobre os mecanismos serão apresentados com base em Little e Rubin (2002), Enders (2010) e Schafer e Graham (2002). Na literatura nacional, boas revisões de literatura sobre a temática estão disponíveis (NUNES, 2007; PÉREZ, 2010; RIBEIRO, 2016).

Os mecanismos foram classificados por Rubin em três grupos: dados faltantes completamente aleatório (*Missing Completely At Random - MCAR*), dados faltantes aleatórios (*Missing At Random - MAR*) e dados faltantes não-aleatórios (*Not Missing At Random - NMAR*). De maneira geral, o mecanismo gerador de dados faltantes é caracterizado pela distribuição condicional da matriz indicadora de dados faltantes $M = \{M_{ij}\}$ dado os dados completos $Y = \{y_{ij}\}$ da forma $f(M|Y, \phi)$, onde ϕ representa parâmetros desconhecidos. Quando M é

independente de Y , ou seja, os dados faltantes não dependem de outras variáveis do conjunto de dados (sejam elas completas ou não), a forma geral fica $f(M|Y, \phi) = f(M|\phi)$ para todo Y e ϕ . Esse caso é definido como dados faltantes completamente aleatório (*Missing Completely At Random - MCAR*). Por outro lado, quando M depende apenas dos valores observados de Y (Y_{obs}), mas não depende dos dados faltantes de Y (Y_{mis}), ou seja, $f(M|Y, \phi) = f(M|Y_{obs}, \phi)$ para todo Y_{mis} e ϕ , os dados são do tipo dados faltantes aleatórios (*Missing At Random - MAR*). Por último, o mecanismo é definido como dados faltantes não-aleatórios (*Not Missing At Random - NMAR*) quando a distribuição de M depende dos valores ausentes na matriz de dados Y , isto é, $f(M|Y, \phi) = f(M|Y_{obs}, Y_{mis}, \phi)$ para todo Y_{mis} e ϕ .

Ao longo do processo de imputação realizado aqui neste trabalho, a suposição assumida é que os dados faltantes foram gerados através de um mecanismo MAR, muito em função do que se observou no Capítulo 3 (a chance de dados faltantes na escolaridade difere dentro das variáveis sociodemográficas). Dessa maneira, a escolha do método para corrigir a incompletitude da escolaridade deve levar em consideração esse pressuposto em conjunto com o de um padrão univariado de falta. Na seção 4.1.2.2 serão apresentados alguns métodos disponíveis na literatura, mas com foco no método escolhido.

4.1.2.2 Métodos de imputação

Os métodos disponíveis para tratar o problema de dados faltantes são amplamente revisados na literatura internacional (COLLINS et al., 2001; LITTLE; RUBIN, 2002; SCHAFER; GRAHAM, 2002; ENDERS, 2010) e nacional (NUNES, 2007; PÉREZ, 2010; RIBEIRO, 2016). Esses métodos foram desenvolvidos, geralmente, com base nas suposições realizadas acerca dos mecanismos que geraram as perdas. Enders (2010) chama atenção para alguns métodos tradicionais que são comumente utilizados, mas que estão cada vez mais em desuso. Alguns deles tiveram origem sob a forte suposição MCAR, ou seja, os dados faltantes não dependem de outras variáveis do conjunto de dados, e são bastante criticados por produzir estimativas tendenciosas dos parâmetros quando essa suposição falhar. Entre esses estão o método de exclusão e a imputação única por alguma medida de tendência central, por exemplo, a média aritmética.

No tocante ao método de exclusão mencionado, a ideia é eliminar do conjunto de dados os casos onde pelo menos uma das variáveis possuem dados ausentes. Depois disso, análises são realizadas como se os dados fossem completos. Além de a exclusão poder ser seletiva caso os dados não sejam MCAR, haverá uma grande perda de casos (redução da amostra) caso se

observe dados faltantes em outras variáveis (ENDERS, 2010). Já os métodos de imputação simples são aqueles que completam o conjunto de dados atribuindo apenas um valor a cada dado faltante e, a partir daí, consideram a base de dados completa. Um exemplo é substituir os valores faltantes pela média aritmética e proceder a análise com o conjunto de dados completo. Apesar de não ocorrer perda de dados, esse método não é recomendado. Ao atribuir um valor constante (a média) aos dados faltantes, as medidas de variabilidade e correlação ficam subestimadas (ENDERS, 2010).

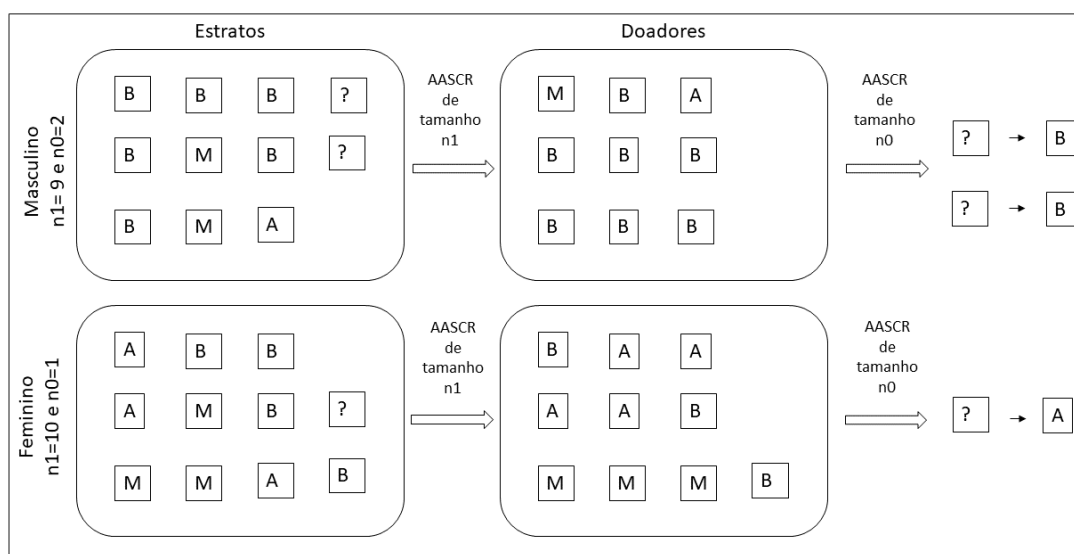
Ainda sobre os métodos que realizam uma única imputação, Enders (2010) chama atenção para Imputação por Regressão (ou média condicional), Imputação regressão estocástica e o método *Hot deck*. Os três utilizam dados de outras variáveis completas em seu processo de imputação e, portanto, deve-se supor dados MAR. Será dada uma maior atenção ao método *Hot deck*, pois a ideia básica do método escolhido para imputar a escolaridade do óbito parte da abordagem original dele. A teoria apresentada a seguir se baseou em Andridge e Little (2010), Allison (2001) e Rubin e Schenker (1986).

A ideia do método *Hot deck* é substituir o dado faltante de uma variável no indivíduo que apresentou ausência (receptor) por um valor dessa mesma variável de um indivíduo com características semelhantes e que a variável a ser imputada foi observada (doador). Basicamente, duas versões estão disponíveis: os métodos *Hot deck* determinísticos e os métodos *Hot Deck* aleatórios. Os primeiros não envolvem aleatoriedade na escolha do doador. Geralmente se utiliza alguma métrica para identificar o vizinho mais próximo para ser o doador. Nos métodos aleatórios, o doador é selecionado aleatoriamente dentro de um conjunto de candidatos a doadores (denominado estratos ou células de imputação) com características similares à unidade com dado faltante.

Há muitas de variações de métodos *Hot Deck* aleatórios, inclusive para tratar dados faltantes em mais de uma variável. Mas é importante lembrar que os pressupostos assumidos acerca dos dados faltantes foram um padrão univariado e o mecanismo MAR. Sendo assim, a variação aqui escolhida é a *ABB* (*Approximate Bayesian Bootstrap*) proposta por Rubin e Schenker (1986). Para ilustrar a ideia, suponha que em um dos estratos possua n_1 casos completos e n_0 casos incompletos na variável a ser imputada Y . Do conjunto dos n_1 casos completos, seleciona-se uma amostra aleatória com reposição de n_1 casos para serem os candidatos a doadores. Em seguida, a partir desta amostra, é selecionada uma amostra aleatória com substituição de n_0 casos. Finalmente, é atribuído os n_0 valores observados de Y aos n_0 casos com dados ausentes em Y . Esse procedimento é repetido em todos os estratos definidos.

A Figura 5 ilustra esse procedimento através de uma situação hipotética. Suponha que em um determinado período e localidade foram observados 22 óbitos. No atestado de óbito só constam duas características do óbito: o sexo (masculino e feminino) e o nível de escolaridade do falecido (B – baixa, M – média e A – alta). Das 22 mortes, 11 são do sexo masculino e 11 do sexo feminino. Considere também que 3 óbitos apresentaram escolaridade faltante, sendo 2 do sexo masculino e 1 do feminino. Suponha também que a variável sexo é uma boa preditora dos dados faltantes de escolaridade e será utilizada no processo de imputação. Sendo assim, o primeiro passo do processo de imputação pelo método *ABB* é a formação de 2 estratos: o dos homens com $n_1 = 9$ e $n_0 = 2$; e o das mulheres com $n_1 = 10$ e $n_0 = 1$. Para os homens, inicialmente extraiu-se uma amostra aleatória simples com reposição (AASCR) de tamanho $n_1 = 9$, onde gerou os possíveis doadores. Na sequência, uma AASCR de tamanho $n_0 = 2$ é selecionada e atribuídas aos receptores. Nesse caso os dois dados faltantes foram preenchidos com baixa escolaridade. O mesmo procedimento foi realizado para as mulheres e uma escolaridade alta foi atribuída ao dado ignorado. De tal forma que, ao final do processo, o conjunto de dados estará completo.

Figura 5: Ilustração do método *ABB* com um caso hipotético.



Fonte: Elaboração própria a partir da ideia do método *ABB* de Rubin e Schenker (1986)

Nota: Escolaridade: A – Alta; M – Média; B - Baixa

Além disso, uma questão central e importante para a aplicação do método *ABB* é a escolha das variáveis utilizadas para criar os estratos com características semelhantes. Dada a importância dessa escolha, ela será tratada no tópico específico 4.1.2.3. Andridge e Little (2010) avaliam que pode ser difícil obter boas correspondências de doadores aos receptores e veem isso como uma fragilidade do método. Contudo, os autores argumentam que em amostras

grandes (como é o caso aqui) é mais provável conseguir bons doadores. Outra exigência do método é que essas variáveis auxiliares sejam categóricas. Para Allison (2001) isso pode ser uma desvantagem do método *ABB*, pois pode limitar o número de variáveis auxiliares utilizadas para formar os estratos. Essa limitação não afeta aqui, porque as variáveis já possuem natureza categórica.

Uma vantagem do método é preservar a natureza da variável imputada, ou seja, no caso da escolaridade categórica com escala ordinal, o valor imputado já é a própria categoria. Outra vantagem que o diferencia de outros métodos (regressão logística ordinal, por exemplo) é que ele evita fortes suposições paramétricas (ANDRIDGE; LITTLE, 2010). Além disso, o método *ABB* é de fácil aplicação e entendimento da mecânica por trás dele.

Também é importante observar que até agora os métodos apresentados se inseriram em um contexto de uma única imputação. Atribuir um único valor ao dado faltante produz estimativas enviesadas por não incorporarem a incerteza da imputação e não é recomendado na literatura (RUBIN, 1987). Por isso, o método *ABB* será utilizado em um contexto de Imputação Múltipla (IM) que será apresentado a seguir.

Ao invés de eliminar os casos incompletos ou realizar uma única imputação aos dados faltantes por algum dos métodos apresentados, a literatura destaca alternativas de métodos mais modernos e superiores a esses procedimentos: Estimação por Máxima Verossimilhança (MV) e Imputação Múltipla (IM). Esses métodos possuem suposições menos restritivas quanto ao mecanismo gerador dos *missing*, assumindo, portanto, que os dados são MAR (ALLISON, 2001; SCHAFER; GRAHAM, 2002). No grupo dos métodos que utilizam estimativas de MV para lidar com dados incompletos, destaca-se o algoritmo EM (*Expectation-Maximization*). Mais detalhes sobre o processo iterativo e da teoria do método podem ser consultados em Dempster et al. (1977). Outra alternativa e que será dada uma maior atenção por ter sido utilizada neste trabalho é a Imputação Múltipla. Os métodos desenvolvidos para lidar com dados faltantes do tipo NMAR não serão tratados aqui, mas podem ser encontrados no capítulo 10 de Enders (2010).

A IM foi proposta por Rubin (1978) para contornar o grande problema dos métodos de imputação simples que é não levar em consideração a variabilidade devida aos dados faltantes. A teoria apresentada aqui sobre IM foi baseada em Rubin (1987) e Rubin e Schenker (1986). A ideia da IM é que para cada dado faltante sejam imputados $m \geq 2$ valores ao invés de apenas um como na imputação simples. A partir daí são gerados m conjunto de dados completos e,

para cada um deles, são realizadas as análises como se os dados fossem completos. Desse modo, a parcela da variabilidade entre as imputações é levada em consideração. Em seguida, faz-se necessário combinar as estimativas geradas através dos m conjuntos de dados completos.

Suponha que o interesse é estimar um parâmetro escalar Q na população. Na ausência de dados faltantes (conjunto de dados completos), seria utilizado um estimador \hat{Q} com variância estimada \hat{V} . Após imputar o conjunto de dados m vezes, são obtidos \hat{Q}_i e \hat{V}_i como valores de \hat{Q} e \hat{V} para o i -ésimo conjunto de dados imputados, $i = 1, \dots, m$. Assim, o estimador IM de Q é a média $\bar{Q} = \frac{1}{m} \sum_{i=1}^m \hat{Q}_i$ e o estimador da variância pelo método IM é dado por:

$$T = W + \left(1 + \frac{1}{m}\right) B,$$

em que $W = 1/m \sum_{i=1}^m \hat{V}_i$ é a variância dentro das imputações e $B = \frac{1}{m-1} \sum_{i=1}^m (\hat{Q}_i - \bar{Q})^2$ é a variância entre as imputações. A proposta é basear inferências sobre Q com base na aproximação da distribuição de $(Q - \bar{Q})/\sqrt{T}$ pela distribuição t de *Student* com v graus de liberdade dado por:

$$v = \left[1 + \left(\frac{m}{m+1}\right) \frac{\hat{W}}{\hat{B}}\right]^2 (m-1).$$

Considerando o cenário abordado aqui neste trabalho, o parâmetro escalar Q a ser estimado é a frequência de óbitos em cada recorte (escolaridade, região e sexo) por grupos etários quinquenais. Após imputar a variável escolaridade do óbito m vezes pelo método *ABB*, serão gerados m conjunto de dados completos. Em seguida, as frequências de óbitos são obtidas em cada conjunto de dados. Por último, é extraída a média e a variância das frequências de óbitos e construídos os intervalos de confiança de 95% em cada recorte considerando a teoria de IM apresentada. É importante ressaltar que como os dados de óbito são de universo, essas contagens são populacionais e não amostrais. Isso significa que a variância dentro das imputações é zero, ou seja, \hat{V} e cada \hat{V}_i são iguais a zero, logo $\hat{W} = 0$. Dessa maneira, os graus de liberdade da distribuição t de *Student* ficam reduzidos a $v = (m-1)$.

Uma questão relevante na IM é a escolha da quantidade de imputações m a serem realizadas. Alguns autores argumentam que m entre 3 e 5 já são suficientes para gerar resultados satisfatórios (RUBIN, 1996; SCHAFER, 1997). Eles se baseiam em um indicador denominado por Rubin de Eficiência Relativa (ER). Esse indicador é função da quantidade de imputações (m) e do percentual de dados ausentes (λ) e é dada por

$$ER = \left(1 + \frac{\lambda}{m}\right)^{-1}.$$

De acordo com os dados observados, o percentual de dados faltantes na variável escolaridade foi de aproximadamente $\lambda = 30\%$. Assim, $m = 3$ e $m = 5$ gerariam uma eficiência relativa satisfatória de 90% e 94%, respectivamente.

Porém, outra parte da literatura parece concordar que é necessário estabelecer um número maior para m (GRAHAM et al., 2007; BODNER, 2008). Eles argumentam que para estimativas pontuais um número pequeno de imputações é razoável, mas podem produzir quantidades como valores-p e intervalos de confiança imprecisos. Essas avaliações se baseiam não apenas no ER, mas em exercícios de simulações para observar o comportamento do poder dos testes de hipóteses (GRAHAM et al., 2007) e da validade dos intervalos de confiança produzidos (BODNER, 2008). Para um percentual de perda de 30%, esses autores sugerem $m = 20$ e $m = 24$, respectivamente.

Dessa maneira, utilizou-se $m = 25$. Decidiu-se utilizar um número maior de imputações para eliminar qualquer dúvida sobre vieses advindos da utilização de um m pequeno. O tempo computacional foi relativamente baixo para todo o processo com $m = 5$ em um teste prévio realizado, de tal modo que considerar um $m = 25$ não seria problema em termos de tempo de processamento computacional.

4.1.2.3 Seleção das variáveis para o método Hot Deck com seleção via *ABB*

Baseado na análise das razões de chances realizadas no Capítulo 3, observou-se evidências de que todas as variáveis do estudo são determinantes para a não declaração da escolaridade do óbito e, portanto, deveriam ser consideradas por métodos de imputação que se baseiam em outras características disponíveis dos óbitos que apresentaram dados faltantes na escolaridade no processo de imputação.

Entretanto, uma das exigências desses métodos é que as variáveis utilizadas para compor estratos homogêneos e, conseqüentemente, os candidatos a doadores da escolaridade aos que não possuem registro (receptores), tenham todos os registros completos. Essa condição está implícita ao supor que os dados são do tipo MAR. Em princípio, essa condição não é verificada, pois ficou evidenciado no Capítulo 3 que, exceto as variáveis Região e Causa básica de morte, todas as outras também apresentaram dados faltantes. Uma alternativa seria imputar também as variáveis auxiliares antes de imputar a escolaridade ou trabalhar com métodos para padrão geral

de dados faltantes (multivariado), mas seria um exercício exaustivo que fugiria do foco desse trabalho.

Outra opção seria excluir os casos incompletos de todas as variáveis auxiliares, mas isso geraria uma perda considerável de casos onde a escolaridade é faltante e teria que ser imputada. As variáveis estado civil e cor/raça apresentaram percentuais consideráveis de casos em que essas variáveis não foram registradas na declaração de óbito conjuntamente com a variável escolaridade, 23,6% e 14,0%, respectivamente. Se esses casos fossem excluídos, o impacto seria uma perda de 37.102 casos, o que representa 11,2% do total de registros de óbito e 37,6% dos casos a serem imputados. Para as variáveis sexo, idade e local de ocorrência esses percentuais foram menos representativos, sendo 0,2%, 3,9% e 0,9%, respectivamente. As variáveis Região e Causa básica de morte não apresentaram dados faltantes. Sendo assim, por essa razão as variáveis estado civil e raça/cor seriam candidatas a não serem utilizadas na imputação, apesar de sua importância demográfica.

Uma questão que merece atenção na seleção desses atributos é a quantidade de doadores nas combinações entre as categorias das variáveis que gerarão os estratos. Considerar todas as variáveis e suas categorias tornam os estratos mais homogêneos e mais assertivos na doação da escolaridade. Contudo, a chance de surgir um receptor com escolaridade faltante, mas que não tenha nenhum candidato a doador aumenta na mesma velocidade em que as combinações das categorias das variáveis aumentam. Se isso ocorrer, a imputação para esses casos fica comprometida. Outra questão é a escassez de doadores em um estrato. Isso pode causar uma sobre-utilização de uma única unidade para imputar os casos (ANDRIDGE; LITTLE, 2010).

Por essas razões, foram analisados três cenários de combinação de variáveis. O primeiro considerou todas as variáveis; o segundo excluiu estado civil e cor/raça; o terceiro excluiu-se estado civil, cor/raça e local de ocorrência. Nos três cenários foram considerados apenas os casos onde as variáveis auxiliares eram completas, ou seja, foram excluídas da base de dados todos os casos onde pelo menos uma variável auxiliar era ausente.

Ao incluir todas as variáveis no primeiro cenário, observou-se um total de 28.000 combinações possíveis entre as categorias das variáveis, Região (5), Sexo (2), Idade (7), Raça/cor (4), Estado civil (4), Local de ocorrência (5), Causa básica de morte (5). Esse montante representa as possíveis formações de estratos ao considerar todas as variáveis e categorias. Sendo assim, observa-se que 778 estratos (2,8%) não possuem nenhum possível doador, ou seja, dentro desses estratos os dados faltantes não seriam imputados. Isso mostra que utilizar

todas as variáveis e categorias inviabiliza a utilização desse tipo de método. Ainda no cenário 1, 23% dos estratos teriam apenas de 1 a 5 doadores. No cenário 2, com a exclusão das variáveis cor/raça e estado civil, seriam formados 1.750 estratos dos quais 31 (1,8%) não teria nenhuma unidade doadora da escolaridade. Estratos com 1 a 5 doadores representaram 21,0% nesse cenário. Por último, no cenário 3, apenas as variáveis Região, sexo, idade e Causa básica de morte foram mantidas, excluindo a variável local de ocorrência (em relação ao cenário 2). Para esse caso, todos os estratos (total de 350 combinações das categorias) apresentaram pelo menos 6 candidatos a doadores. A Tabela 4 apresenta os resultados dessa breve simulação.

Tabela 4: Frequência absoluta e relativa de estratos segundo a condição da quantidade de doadores disponíveis para imputar escolaridade do óbito.

Condição dos estratos	Cenário 1		Cenário 2		Cenário 3	
	freq.	%	freq.	%	freq.	%
Sem doador	778	2,8%	31	1,8%	Não houve estrato com menos de 6 doadores	
Com 1 doador	3019	10,8%	121	6,9%		
Com 2 a 5 doadores	3405	12,2%	246	14,1%		
Com mais de 5 doadores	20798	74,3%	1352	77,3%		
Total de combinações das categorias	28000	100,0%	1750	100,0%	350	100,0%

Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010.

Variáveis do Cenário 1: Região, Sexo, Idade, Cor/raça, Estado civil, Local de ocorrência e Causa básica de morte

Variáveis do Cenário 2: Região, Sexo, Idade, Local de ocorrência e Causa básica de morte

Variáveis do Cenário 3: Região, Sexo, Idade e Causa básica de morte

Diante do que foi observado nessa simulação e na análise de perda de casos, as variáveis que serão utilizadas para compor os estratos são: Região, Sexo, Idade e Causa básica de morte. Com isso, foram excluídos da base de dados final os casos onde pelo menos sexo ou idade eram faltantes, saindo de um total de 329.815 óbitos para 325.829, uma perda de apenas 3.986 (1,2%) registros de óbito. Vale salientar que decidiu-se utilizar a variável escolaridade a ser imputada com a categorização original (Nenhuma, 1 a 3 anos, 4 a 7 anos, 8 a 11 anos, 12 anos ou mais, Ignorado), o que torna as doações da escolaridade aos receptores mais fidedignas.

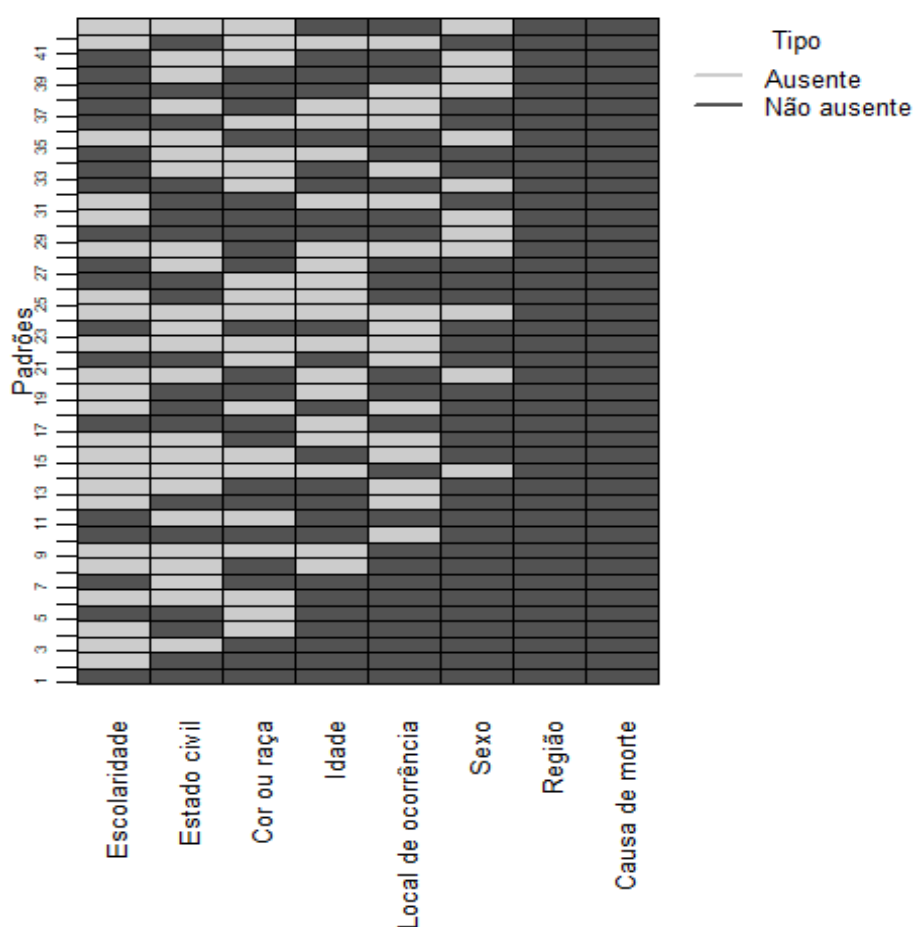
4.2 Resultados da imputação

Antes de apresentar os resultados da imputação, é apresentada uma análise dos padrões de dados faltantes presentes no conjunto de dados (Figura 6). Esse gráfico é útil para ajudar a encontrar indícios sobre o padrão das ausências. O eixo y do gráfico representa cada padrão identificado e no eixo x cada variável do estudo. Ao todo, identificou-se 42 padrões distintos de dados faltantes. Para exemplificar, os registros de óbitos que não apresentaram dados faltantes em nenhuma das variáveis se encaixam no padrão 1. Já no padrão 24, estão os óbitos

que apresentaram ausências concomitantes no registro das variáveis sexo, local de ocorrência, idade, cor ou raça, estado civil e escolaridade.

Fica evidente que o padrão dos dados faltantes é do tipo padrão geral, ou seja, há ocorrência de dados faltantes em mais de uma variável no conjunto de dados (Figura 6), corroborando o que foi evidenciado no Capítulo 3. Porém, conforme justificado e definido na seção 4.1.2.1, a suposição adotada é que os dados faltantes possuem um padrão univariado e foram gerados pelo mecanismo MAR.

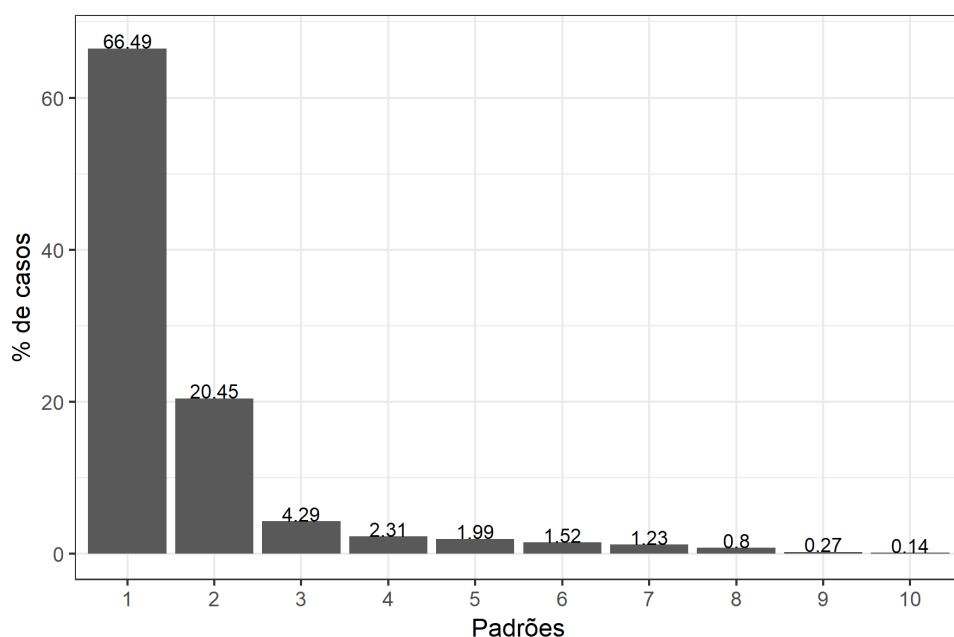
Figura 6: Padrões dos dados faltantes nos registros de óbitos do SIM no Brasil em 2010, segundo variáveis sociodemográficas.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010.

Ainda sobre os padrões, observa-se que, do total de óbitos, a maioria (66,49%) dos registros se enquadram no padrão 1 (nenhuma variável é ausente), seguido de 20,45% dos casos classificados no padrão 2 (apenas a escolaridade é faltante). Os percentuais de casos que apresentaram ausência em outras variáveis em conjunto com ausência na variável escolaridade são pouco expressivos (Figura 7). Esse resultado evidencia a razoabilidade de se assumir um padrão univariado, em que apenas a escolaridade possui dados faltantes.

Figura 7: Distribuição dos óbitos segundo os 10 padrões mais frequentes no conjunto de dados.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010.

Nota: Para identificar os padrões, consultar Figura 6.

No que se refere à imputação, a ideia aqui é apresentar os resultados obtidos com a aplicação do método ABB e imputação múltipla. O produto final da imputação são as estimativas das médias dos óbitos por idade e seus respectivos intervalos de confiança em cada recorte, em um total de 30 tabelas (disponíveis no [Repositório da dissertação - scripts em R e resultados](#)). Outro resultado apresentado é a comparação da distribuição proporcional média por idade obtida com o processo de imputação com a distribuição proporcional observada sem imputação (excluindo os dados faltantes) também para cada recorte. Essa comparação tem o objetivo de verificar se o padrão dos óbitos por idade muda após o processo de imputação. Também são analisadas essas diferenças segundo a escolaridade, com o propósito de identificar qual categoria de escolaridade obteve maior ganho de casos com a imputação. Os resultados dessas comparações são apresentados apenas para o Brasil. As diferenças por região estão disponíveis no [Repositório da dissertação - scripts em R e resultados](#).

Na Tabela 5, são apresentadas as distribuições proporcionais por idade em cada grupo de escolaridade e sexo. Para o Brasil, observa-se que as diferenças entre as proporções de óbitos por idade são mínimas, onde a diferença máxima encontrada nos homens foi de 0,25% e nas mulheres 0,10%. Isso é uma forte evidência de que as distribuições proporcionais são iguais.

Tabela 5: Comparação entre a distribuição de óbitos por idade com e sem imputação segundo sexo e escolaridade, Brasil, 2010.

Sexo	Idade	Escolaridade								
		Baixa			Média			Alta		
		Com imputação	Sem imputação	Diferença	Com imputação	Sem imputação	Diferença	Com imputação	Sem imputação	Diferença
Masculino	25 a 30	9,24%	9,40%	0,16%	13,95%	14,20%	0,25%	10,62%	10,66%	0,04%
	30 a 35	9,82%	9,73%	-0,08%	11,24%	11,18%	-0,07%	9,83%	9,76%	-0,07%
	35 a 40	10,55%	10,55%	0,00%	10,43%	10,40%	-0,03%	8,63%	8,67%	0,03%
	40 a 45	12,92%	12,78%	-0,13%	12,05%	11,89%	-0,15%	10,55%	10,46%	-0,09%
	45 a 50	15,86%	15,81%	-0,05%	15,43%	15,37%	-0,06%	14,72%	14,68%	-0,05%
	50 a 55	19,22%	19,22%	0,00%	18,30%	18,27%	-0,02%	20,72%	20,78%	0,06%
	55 a 60	22,40%	22,50%	0,10%	18,60%	18,68%	0,08%	24,93%	25,00%	0,07%
	Total	100,00%	100,00%		100,00%	100,00%		100,00%	100,00%	
Feminino	25 a 30	4,38%	4,40%	0,02%	9,22%	9,24%	0,03%	8,69%	8,76%	0,06%
	30 a 35	6,19%	6,21%	0,02%	9,93%	9,99%	0,07%	9,54%	9,54%	-0,01%
	35 a 40	8,08%	8,09%	0,01%	10,83%	10,75%	-0,08%	9,63%	9,61%	-0,02%
	40 a 45	11,99%	12,03%	0,04%	13,19%	13,28%	0,09%	12,12%	12,21%	0,09%
	45 a 50	17,33%	17,36%	0,03%	17,33%	17,36%	0,03%	17,23%	17,26%	0,02%
	50 a 55	22,83%	22,80%	-0,03%	20,22%	20,15%	-0,08%	19,85%	19,78%	-0,07%
	55 a 60	29,21%	29,11%	-0,10%	19,28%	19,23%	-0,05%	22,92%	22,84%	-0,08%
	Total	100,00%	100,00%		100,00%	100,00%		100,00%	100,00%	

Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010.

Quando se calcula a distribuição percentual em relação ao total da variável escolaridade, observa-se que os ganhos nas participações de cada categoria de escolaridade foram mínimos. No sexo masculino, o maior ganho de participação após imputação foi na escolaridade média (0,10%), enquanto no sexo feminino praticamente não houve diferença (Tabela 6).

Tabela 6: Comparação entre a distribuição de óbitos com e sem imputação segundo sexo e escolaridade, Brasil, 2010.

Sexo	Escolaridade	Com imputação	Sem imputação	Diferença
Masculino	Baixa	71,50%	71,61%	-0,11%
	Média	20,29%	20,19%	0,10%
	Alta	8,21%	8,20%	0,01%
	Total	100,00%	100,00%	
Feminino	Baixa	69,14%	69,12%	0,02%
	Média	19,76%	19,76%	0,00%
	Alta	11,09%	11,12%	-0,03%
	Total	100,00%	100,00%	

Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010.

4.3 Discussão

Os resultados das imputações trazem um panorama interessante quando foram comparadas as distribuições proporcionais por idade em cada grupo de escolaridade e sexo. As evidências mostraram que, ao imputar, não houve distorção no padrão de mortes. Isso significa que o padrão observado excluindo os dados faltantes de escolaridade podem ser utilizados em estudos de mortalidade sem grandes problemas, mas seu nível não. O mesmo foi observado na composição dos óbitos por escolaridade segundo o sexo.

É importante ressaltar que o objetivo aqui não é esgotar as possibilidades metodológicas para imputar a escolaridade do óbito adulto em 2010, mas sim escolher um método com critérios e suposições pré-estabelecidas que sejam razoáveis para corrigir o problema. Outros métodos de imputação podem ser utilizados em trabalhos futuros para comparação das estimativas aqui realizadas. A literatura tem sugerido utilizar mais de um método e compará-los segundo alguns critérios de avaliação (COLLINS et al., 2001). Mas essa prática demandaria um investimento de tempo muito grande que poderia afetar nas análises sobre os diferenciais de mortalidade que é o foco do estudo. Outros métodos utilizados no contexto de estudos sobre diferenciais de mortalidade foram utilizados por Ribeiro (2016) e Pérez (2010).

Uma limitação na aplicação da metodologia é que foi assumido que as variáveis sexo e idade possuem dados MCAR e o método da exclusão foi utilizado para tomá-las como uma variável auxiliar para compor os estratos e, assim, imputar a escolaridade. Porém, a literatura sugere que essa prática é razoável quando o percentual de dados faltantes nas variáveis é pequeno (HARRELL, 2001; ENDERS, 2010). Uma alternativa utilizada, por exemplo pelo *US Bureau of the Census*, é realizar imputações sequenciais e individuais das variáveis auxiliares, permitindo utilizá-las com todos os casos completos na composição dos estratos no método Hot Deck (ANDRIDGE; LITTLE, 2010). Outra questão que merece atenção é a relação entre os dados faltantes das variáveis sociodemográficas e os da escolaridade observada no Capítulo 3. Embora na formação dos estratos apenas sexo e idade tenham apresentado ausências de registro, a relação entre as perdas pode ser um indício de que os dados possam ter sido gerados pelo mecanismo NMAR e produzir estimativas enviesadas aqui. Ainda sobre as variáveis utilizadas para compor os estratos, é possível que variáveis de confundimento que não foram mensuradas estejam relacionadas com a ausência da variável escolaridade.

No que diz respeito à utilização da metodologia clássica de imputação múltipla desenvolvida por Rubin (1987) em um contexto de dados de universo, como é o caso aqui,

deve-se alertar para algumas considerações. Essas técnicas de imputação múltipla podem apresentar problemas quando a parcela da variância devida ao erro de amostragem é zero em teste de hipóteses multivariados (KINNEY, 2012). Embora não tenha sido verificado, as questões levantadas por esse autor podem ter influência nas estimativas das variâncias e, conseqüentemente, na construção de intervalos de confiança para um escalar Q. Ainda sobre a metodologia de IM, outros autores sugerem alternativas para o cálculo da variância para situações onde o estimador proposto por Rubin é inconsistente (FAY, 1996). Outra alternativa é sugerida para o cálculo da variância quando a amostra é pequena ou moderada (não é o caso aqui) na utilização do método ABB na IM (KIM, 2002).

5 DIFERENCIAIS DE MORTALIDADE ADULTA NO BRASIL, SEGUNDO ESCOLARIDADE

O objetivo deste capítulo é apresentar os principais resultados sobre as diferenças nos riscos e nas probabilidades de morte segundo o nível de escolaridade do óbito adulto em 2010. Antes de prosseguir com as estimativas de mortalidade, ainda se faz necessário alguns ajustes no numerador e denominador das taxas de mortalidade.

Por isso, pretende-se apresentar também neste capítulo as estratégias utilizadas para correção de sub-registro, a compatibilização entre as variáveis de escolaridade do SIM e do Censo 2010, o cálculo das medidas de exposição e das medidas de mortalidade utilizadas.

5.1 Metodologia

5.1.1 Fonte de dados

Duas fontes de dados são utilizadas para se produzir as estimativas de mortalidade. A primeira, dos óbitos, são os microdados do SIM e já foi amplamente discutida nos Capítulos 3 e 4. A segunda fonte de dados está relacionada aos dados populacionais, em que se obteve os quantitativos populacionais por escolaridade segundo grandes regiões, sexo e idade simples do Censo Demográfico 2010. Eles são úteis para o cálculo das medidas de exposição (pessoas-ano) para 2010 utilizadas nos denominadores das taxas de mortalidade. Na Tabela 7, apresenta-se a descrição e categorização das variáveis do Censo 2010.

Tabela 7: Variáveis sociodemográficas do censo demográfico 2010 utilizadas.

Código da variável	Nome da variável	Descrição	Categorias
V1001	Região	Região geográfica	1: Região norte (uf=11 a 17) 2: Região nordeste (uf=21 a 29) 3: Região sudeste (uf=31 a 33 e 35) 4: Região sul (uf=41 a 43) 5: Região Centro-oeste (uf=50 a 53)
V0601	Sexo	Sexo da pessoa recenseada	1: Masculino 2: Feminino
V6036	Idade	Idade da pessoa em anos completos na data de referência da pesquisa	Variável discreta em anos completos
V6400	E escolaridade	Nível de instrução da pessoa recenseada	1: Sem instrução e fundamental incompleto 2: Fundamental completo e médio incompleto 3: Médio completo e superior incompleto 4: Superior completo 5: Não determinado

Fonte: IBGE - Censo demográfico 2010

A utilização da variável idade no formato de idade simples de 0 a 99 anos (em anos completos) se justifica porque é a configuração requerida como *input* para o método de cálculo da medida de exposição sugerida por Gonzaga e Schmertmann (2016). Porém, ao final do processo de estimação das medidas de exposição por idade simples, agregou-se as quantidades populacionais em grupos quinquenais e, para as estimativas de mortalidade, apenas as idades adultas foram utilizadas (25 a 29 até 55 a 59 anos).

Outro detalhe importante se refere ao uso em conjunto dos quantitativos de óbitos do SIM e populacionais do Censo 2010 segundo a variável escolaridade, sendo necessário um processo de compatibilização. Isso porque a variável escolaridade é mensurada de maneira distinta nas duas fontes. Nos microdados do SIM, a escolaridade é disponibilizada em termos de faixa de anos de estudo. Por outro lado, esse atributo se apresenta em função do nível de instrução no Censo 2010.

As categorias da escolaridade do SIM “Nenhuma”, “1 a 3 anos de estudo” e “4 a 7 anos de estudo” são correspondentes ao nível “Sem instrução / Fundamental incompleto” da variável escolaridade do censo. Para o nível “8 a 11 anos de estudo” do SIM é atribuído o nível do censo “Fundamental completo / Médio incompleto”. Por último, a categoria “12 anos de estudo e +” é compatível com “Médio completo/Superior Incompleto” e “Superior completo”. E o “Ignorado” do SIM é o “Não determinado” do censo. Para facilitar a apresentação dos resultados, atribuiu-se as classificações baixa, média e alta para a escolaridade, conforme Tabela 8.

Tabela 8: Compatibilização entre a variável “Anos de estudo” do SIM 2010 e “nível de instrução” do Censo 2010.

Escolaridade - SIM 2010	Escolaridade - CENSO 2010	Categorização utilizada para escolaridade
Nenhuma	Sem instrução/Fundamental Incompleto	
1 a 3 anos de estudo	Sem instrução/Fundamental Incompleto	Baixa
4 a 7 anos de estudo	Sem instrução/Fundamental Incompleto	
8 a 11 anos de estudo	Fundamental completo/Médio Incompleto	Média
12 anos de estudo e +	Médio completo/Superior Incompleto/Superior completo	Alta
Ignorado	Não determinado	Dado faltante

Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010 e IBGE - Censo demográfico 2010

Deve-se levar em consideração também que a variável “nível de instrução” do censo de 2010 possui dados faltantes. Entretanto, observa-se que os percentuais de dados ignorados foram modestos, não ultrapassando 0,59% de representatividade tanto no sexo masculino quanto para o feminino em todas as regiões brasileiras. Esses percentuais reduzem com a idade

em todos os recortes. Mais detalhes podem ser vistos nas Tabelas A3 no apêndice. Como os percentuais são pequenos, esses indivíduos foram excluídos do estudo.

5.1.2 Ajustes nos dados

No tocante à correção de sub-registro de óbitos, não é objetivo desta dissertação estimar os graus de cobertura por escolaridade, região e sexo. Ao invés disso, a estratégia foi utilizar as estimativas dos fatores de correção para a população geral por UF propostas por Queiroz et al. (2017) e corrigir apenas os óbitos com baixa escolaridade (≤ 7 anos de estudo). Os autores estimaram os fatores de correção por UF e por sexo para os óbitos de 1980/1991, 1991/2000 e 2000/2010 utilizando os métodos GGB - *General Growth Balance* desenvolvido por Hill (1987) e SEG adj.- Gerações Extintas Sintético Ajustado proposto por Hill et al. (2009).

Aqui optou-se pelas estimativas obtidas por meio do método SEG adj. por Queiroz et al. (2017). A ideia do método SEG adj. é combinar o método GGB de Hill (1987) com o método SEG de Bennett e Horiuchi (1981). Caso haja erro de cobertura nos censos, se corrige com as estimativas do GGB. Depois disso, aplica-se o SEG para corrigir os óbitos. O SEG adj. é recomendado quando não se possuem informações em relação à variação do grau de cobertura por idade, bem como sobre como a população é afetada pela migração. Para esses casos, após algumas simulações de erros realizadas, Hill et al. (2009) verificaram que o método SEG adj. foi o que apresentou o menor erro quadrático médio comparado à aplicação dos métodos separadamente.

Para obter cada fator de correção para região e sexo, calculou-se a média ponderada dos fatores de correção estimados para as UFs em cada região e sexo, onde o peso é o número de óbitos de 15 a 59 anos na UF. Assim, o fator de correção ponderado é dado por

$$FC_{j,w} = \sum_{i=1}^k \left(\frac{FC_{i,j,w} * O_{i,j,w}}{O_{j,w}} \right),$$

onde:

$FC_{j,w}$ é o fator de correção ponderado para a região j = Norte, Nordeste, Sul, Sudeste e Centro-oeste e sexo w = Masculino e Feminino;

$FC_{i,j,w}$ é o fator de correção para as k unidades da federação na região j e sexo w ;

$O_{i,j,w}$ é o quantitativo de óbitos em cada unidade da federação na região j e sexo w ;

$O_{j,w}$ é o quantitativo de óbitos na região j e sexo w .

Na literatura não há registro de trabalhos que estimaram o grau de cobertura dos óbitos adultos por escolaridade, sexo e grandes regiões. No trabalho de Silva (2014) e Silva et al. (2016) para dados de óbito do censo 2010, eles corrigiram os óbitos por escolaridade, mas estimando fatores de correção da população geral por escolaridade para corrigir apenas os óbitos onde a pessoa de referência possuía os menores níveis de escolaridade. O método utilizado foi o SEG Adj. A suposição é que a concentração de sub-registros ocorre nos menos instruídos, pois um nível baixo de escolaridade de uma população está relacionado a um baixo desenvolvimento do local de residência dessa população. Consequentemente, pior tende a ser a cobertura dos registros de óbitos (HILL et al., 2009; SILVA, 2014; SILVA et al., 2016).

Além disso, essa forma de correção foi justificada por Silva (2014) e Silva et al. (2016) pelo fato de ter se observado que as taxas de crescimento populacionais se comportaram de maneira inadequadas (muito elevadas em algumas idades) ao aplicar o método para todas as categorias de escolaridade. Para eles, isso é um indicativo da quebra do pressuposto de populações fechadas, já que se observou uma mobilidade considerável entre os níveis de escolaridade de 2000 para 2010. Desta forma, a mesma estratégia de correção foi utilizada neste trabalho.

No que se refere ao denominador utilizado nas estimativas das taxas de mortalidade, calculou-se a medida de exposição sugerida por Gonzaga e Schmertmann (2016). A ideia dessa medida foi pensada em um contexto de estimativas de mortalidade por idade simples de 0 a 99 anos e para três anos, sendo o ano censitário, o anterior e o posterior (por exemplo, 2009,2010,2011). Entretanto, a estimativa de pessoas-ano específica por idade e sexo para o período de três anos é obtida a partir apenas da população do censo 2010. Sob o ponto de vista do presente estudo, o interesse é calcular as medidas de exposição por grupo quinquenal segundo escolaridade, região e sexo para o ano de 2010. Como o resultado da função é uma estimativa de exposição para três anos, dividiu-se por 3 para obter apenas para 2010, assumindo homogeneidade nos 3 anos de exposição, ou seja, a exposição em 2009 é igual a de 2010 que é igual a de 2011. Ao final, agrupou-se as estimativas de exposição em grupos quinquenais. Mais detalhes teóricos sobre o método e acesso à função original do R desenvolvida pelos autores podem ser obtidos em <http://topals-mortality.schmert.net/>.

A escolha desse método para calcular a medida de exposição se justifica por duas razões. A primeira é que as taxas de crescimento da população por escolaridade entre 2000 e 2010 possuem distorções (SILVA, 2014). A segunda é porque as categorias de escolaridade dos censos 2000 e 2010 não são comparáveis e seria necessário outra compatibilização (SILVA,

2014) . Por esses motivos, optou-se por não utilizar o cálculo tradicional de pessoas-ano pelas taxas de crescimento entre 2000 e 2010, trazendo a população para o meio do ano de 2010 que é 01/agosto/2010. Uma vantagem do método proposto por Gonzaga e Schmertmann (2016) é que utiliza apenas dados de 2010.

Após todos os procedimentos de correção, inclusive a imputação dos dados faltantes já tratada no capítulo 4, os resultados são apresentados em termos das taxas específicas de mortalidade (TEM) e de uma medida resumo de mortalidade que é a probabilidade de morte dos 25 aos 59 anos (35q25) obtidas pelas funções clássicas das tabelas de sobrevivência (ORTEGA, 1987; PRESTON et al., 2001). Além disso, são calculadas as razões de risco ($RR = TEM \text{ nos óbitos de baixa escolaridade} / TEM \text{ nos óbitos de alta escolaridade}$) e as razões de probabilidade ($RP = 35q25 \text{ nos óbitos de baixa escolaridade} / 35q25 \text{ nos óbitos de alta escolaridade}$), medidas essas que facilitam a visualização dos diferenciais.

Todos os resultados são apresentados em termos médios e seus respectivos intervalos de confiança obtidos na imputação múltipla do numerador das taxas. Isso não quer dizer que esses limites representam os intervalos de confiança para cada medida de mortalidade, mas sim a obtenção das medidas de morte caso os limites inferiores e superiores das quantidades de óbitos da imputação múltipla fossem utilizados. Outro detalhe é que são apresentados dois cenários de resultados conforme descritos nos objetivos, um com correção de sub-registro (apenas para a categoria com menor escolaridade) e outro sem essa correção. As taxas específicas de mortalidade por 100.000 habitantes segundo a escolaridade (Tabelas A4 e A5) e os gráficos das probabilidades de morte (35q25) e das razões de probabilidade (RP) (Figuras A1 a A6) estão disponíveis no Apêndice e no [Repositório da dissertação - scripts em R e resultados](#).

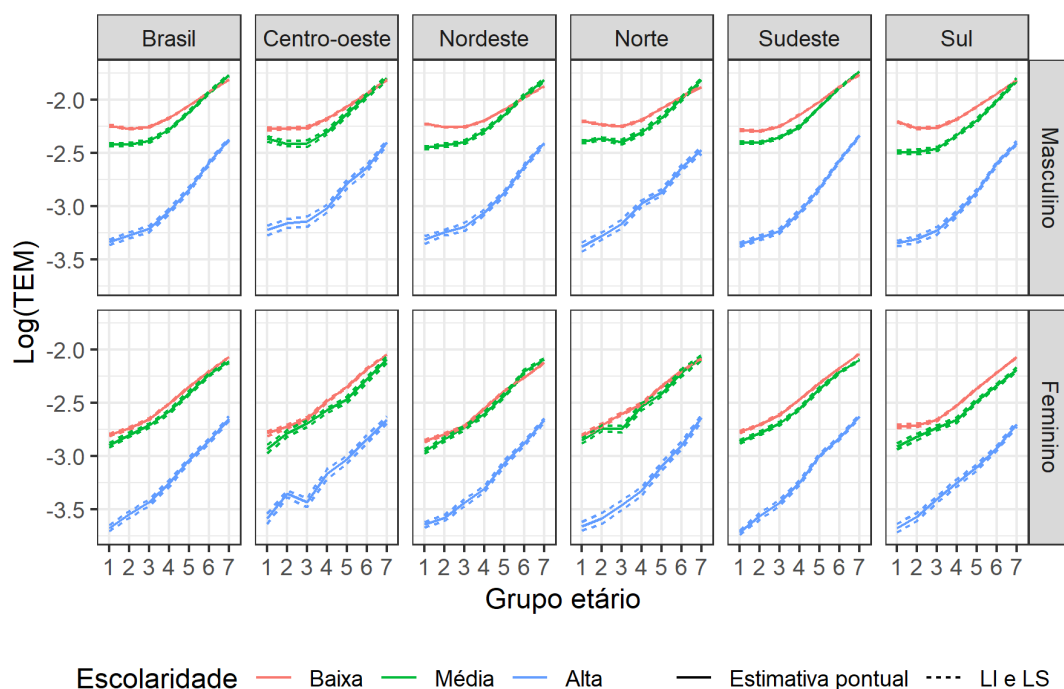
5.2 Resultados

Com relação aos diferenciais de mortalidade por nível de escolaridade, observa-se claramente que quanto maior a escolaridade menor são os riscos de mortalidade para todos os recortes (conforme indica a literatura), tanto no cenário em que foi realizado correção de sub-registro, quanto sem essa correção. Entretanto, algumas ressalvas devem ser colocadas. Isso fica evidente quando se compara a curva de alta escolaridade com as demais. Por outro lado, quando se compara a média escolaridade e a baixa escolaridade, essa relação inversa não está bem definida em algumas categorias, principalmente nos homens de idade mais avançada.

Por exemplo, analisando o Brasil como um todo no cenário com correção de sub-registro (Figura 8), o risco de morte foi igual para os homens que possuíam média escolaridade comparado aos que tinham baixa escolaridade nas idades de 50 a 54 anos e maior nas idades de 55 a 59 anos. Na região Nordeste, o risco de morte para esses dois últimos grupos etários foi maior na escolaridade média com relação à escolaridade mais baixa, tanto para homens quanto para mulheres. Nas demais regiões esse *crossover* é verificado apenas no grupo de 55 a 59 anos e no sexo feminino.

É importante observar que, no cenário em que não se realizou correção de sub-registro (Figura 9), esse efeito se agrava ainda mais, pois assumiu-se que as sub-notificações de óbitos se concentram apenas em quem tem baixa escolaridade. As regiões Norte e Nordeste são as mais afetadas, já que o grau de cobertura é menor nessas regiões. O caso que chama mais atenção é o da região Norte para o sexo feminino. Exceto no grupo de 45 a 49 anos, em todos os grupos etários o risco de morte na média escolaridade foi maior do que na baixa escolaridade. As Figuras 8 e 9 fornecem uma visão geral sobre os diferenciais de mortalidade por escolaridade em cada região e sexo por idade para os dois cenários com e sem correção de sub-registro, respectivamente.

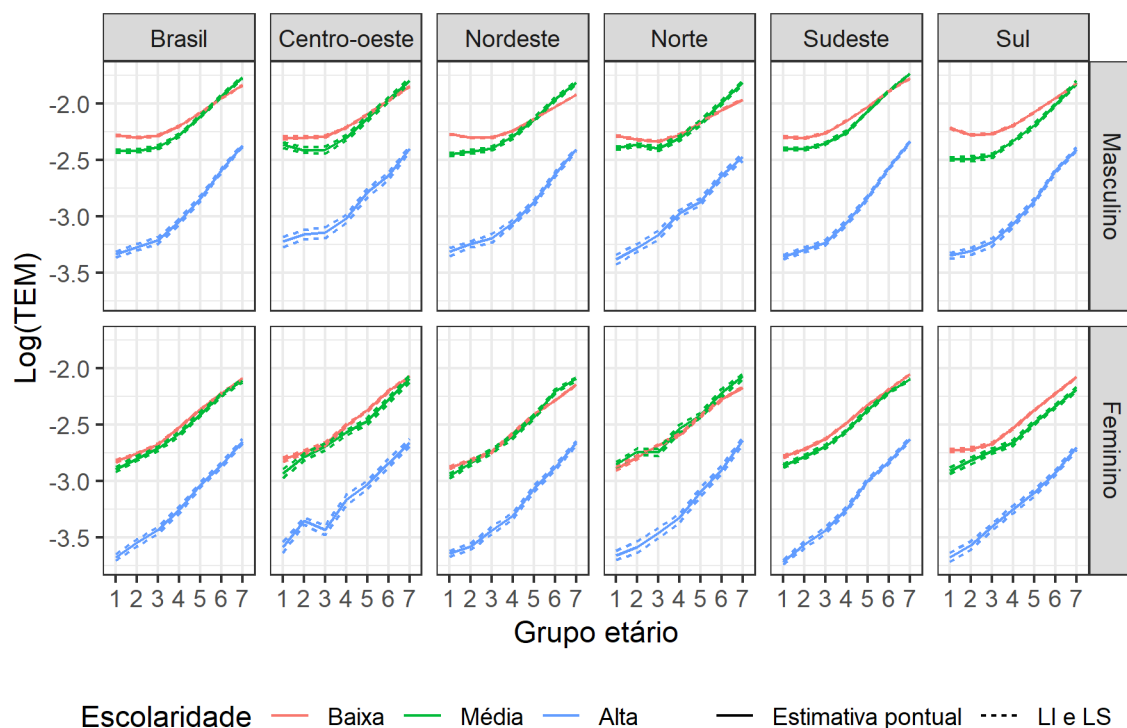
Figura 8: Logaritmo das taxas específicas de mortalidade por nível de escolaridade segundo região e sexo no cenário com correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: Grupo etário = 1 – 25 a 29 anos, 2 – 30 a 34 anos, 3 – 35 a 39 anos, 4 – 40 a 44 anos, 5 – 45 a 49 anos, 6 – 50 a 54 anos, 7 – 55 a 59 anos.

Figura 9: Logaritmo das taxas específicas de mortalidade por nível de escolaridade segundo região e sexo no cenário sem correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: Grupo etário = 1 – 25 a 29 anos, 2 – 30 a 34 anos, 3 – 35 a 39 anos, 4 – 40 a 44 anos, 5 – 45 a 49 anos, 6 – 50 a 54 anos, 7 – 55 a 59 anos.

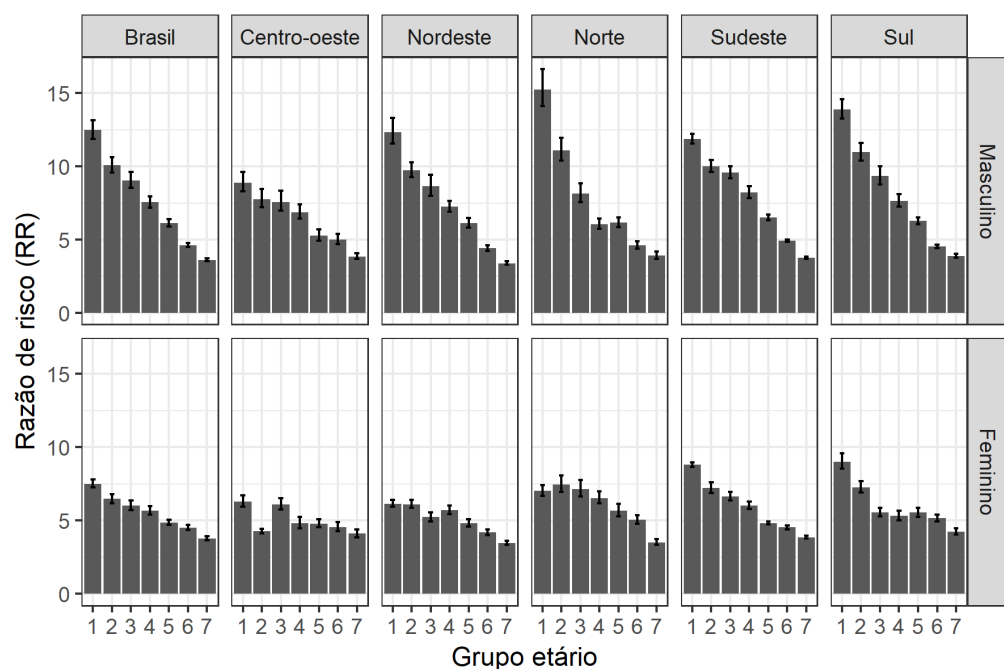
Quando se analisa esses gradientes na mortalidade por educação utilizando uma medida resumo (a probabilidade de morte de 25 a 59 anos - 35q25), a relação inversa entre mortalidade e escolaridade é observada em todos os recortes geográficos e sexo, exceto para as mulheres da região Nordeste (cenário com correção de sub-registro). As nordestinas adultas com baixa escolaridade possuem probabilidade de morte estatisticamente igual a daquelas que possuem média escolaridade. Caso as sub-enumerações dos óbitos não fossem corrigidas, essa relação inversa considerando a baixa e média escolaridade também não seria observada entre homens e mulheres da região Norte e entre as mulheres da região Nordeste (ver Apêndice - Figuras A1 e A2).

No que se refere ao comportamento das razões de risco por idade, observa-se que eles diminuem à medida que a idade aumenta em todas as regiões e para ambos os sexos nos dois cenários de análise. Uma exceção é o caso das mulheres da região Centro-oeste, em que o padrão da magnitude dos diferenciais se mantém constantes a partir dos 40 anos. Observou-se também que essa redução dos diferenciais por idade é mais acentuada entre os homens nos dois cenários analisados. No Brasil (cenário com correção de sub-registro), o risco de morte nos

homens menos escolarizados de 25 a 29 anos é 12,47 [11,88 – 13,14] vezes o risco nos mais escolarizados. Na idade mais avançada (55 a 59 anos), essa razão cai para 3,61 [3,52 – 3,71]. Para as mulheres menos escolarizadas de 25 a 29 anos o risco de morte nas menos escolarizadas é 7,51 [7,26 – 7,79] vezes o risco nas mais escolarizadas. Essa razão de risco é de 3,78 [3,65 – 3,92] nas mulheres em idade mais avançada aqui observadas (55 a 59) (Figura 10). No cenário sem correção de sub-registro, observa-se o mesmo comportamento no padrão dos gradientes por idade, mas em magnitudes ligeiramente menores (Figura 11).

Esses resultados são consistentes com o que foi observado nos Estados Unidos (KITAGAWA; HAUSER, 1973; HUMMER; LARISCY, 2011; MONTEZ et al., 2012), em alguns países da Europa (KUNST et al., 2004) e em alguns da América Latina (MANZELLI, 2014; SANDOVAL; TURRA, 2015). No Brasil, mais especificamente em São Paulo, essa redução por idade não foi encontrada nos óbitos de 25 a 59 anos do censo de 2010 (RIBEIRO, 2016). Embora com recortes de ano, idade e geográficos diferentes, os resultados aqui encontrados corroboram os obtidos por Turra et al. (2018), em que os diferenciais reduzem com a idade e apresentam comportamento mais bem definido para os homens. A hipótese mais comum na literatura é a de seletividade da mortalidade, em que pessoas expostas a baixas condições socioeconômicas e que sobrevivem, passam a ter uma condição de saúde equiparada à de pessoas expostas a altas condições socioeconômicas (LAUDERDALE, 2001). Outra questão que é apontada na literatura é em relação aos efeitos de período e de coorte nessa análise por idade, pois o efeito da escolaridade pode ser menor em coortes mais velhas (LAUDERDALE, 2001). Além disso, existe a possibilidade de erro na declaração de idade da população, principalmente entre os adultos de idade mais avançada (PRESTON et al., 1999). Isso afetaria os dados de população do denominador das taxas de mortalidade.

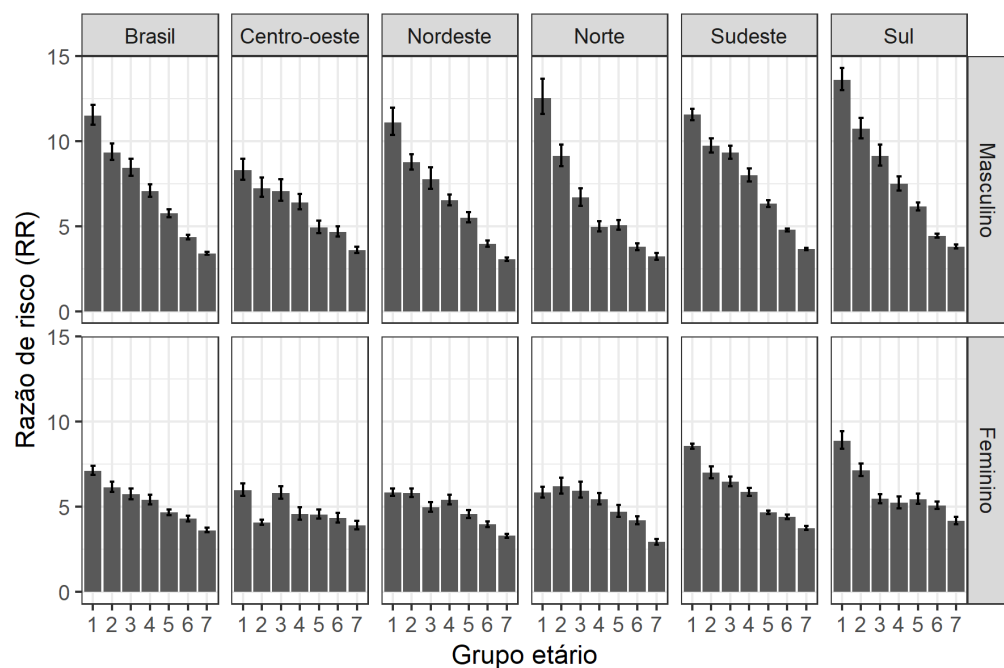
Figura 10: Razão de risco por idade segundo o sexo e grandes regiões no cenário com correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: Grupo etário = 1 – 25 a 29 anos, 2 – 30 a 34 anos, 3 – 35 a 39 anos, 4 – 40 a 44 anos, 5 – 45 a 49 anos, 6 – 50 a 54 anos, 7 – 55 a 59 anos; RR = (Taxa de mortalidade da baixa escolaridade) / (Taxa de mortalidade da alta escolaridade).

Figura 11: Razão de risco por idade segundo o sexo e grandes regiões no cenário sem correção de sub-registro, 2010.

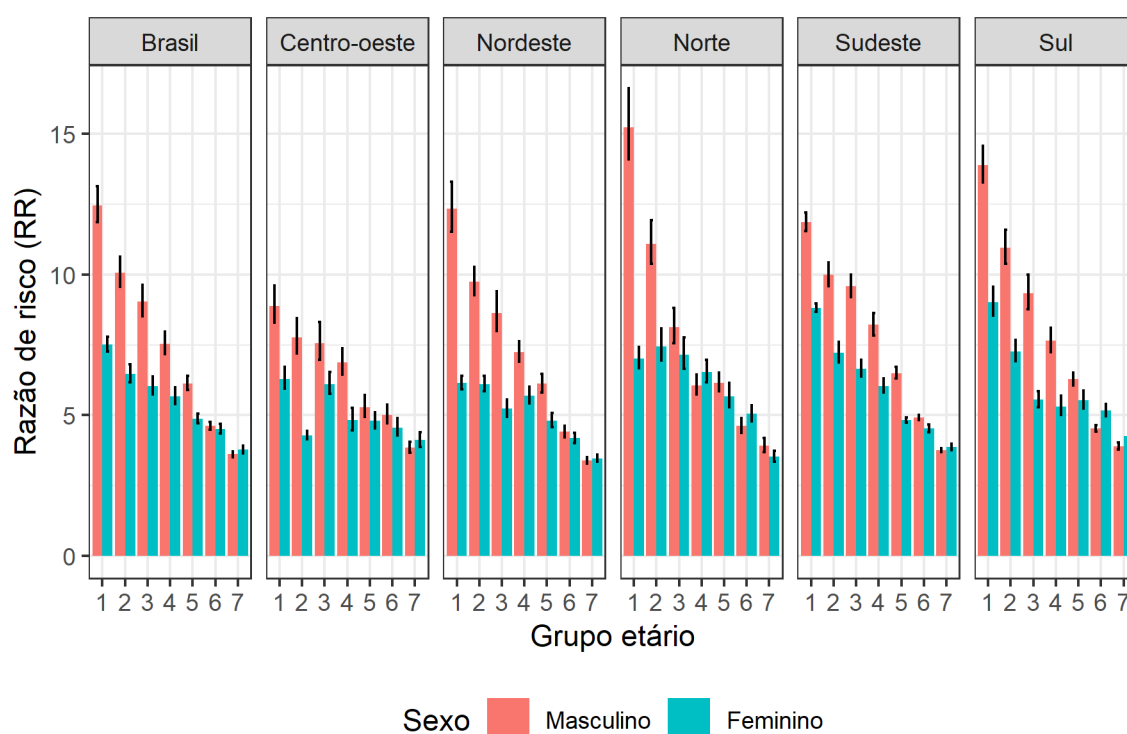


Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: Grupo etário = 1 – 25 a 29 anos, 2 – 30 a 34 anos, 3 – 35 a 39 anos, 4 – 40 a 44 anos, 5 – 45 a 49 anos, 6 – 50 a 54 anos, 7 – 55 a 59 anos. RR = (Taxa de mortalidade da baixa escolaridade) / (Taxa de mortalidade da alta escolaridade).

De modo geral, os diferenciais de mortalidade por escolaridade são maiores entre os homens para o Brasil e regiões, principalmente nos adultos mais jovens nos cenários com e sem correção de sub-enumeração (Figuras 12 e 13). Como a redução dos diferenciais é mais acentuada nos homens quando se aumenta a idade, os diferenciais de gênero tendem a diminuir com a idade. Uma explicação é que a educação como capital humano pode ter efeito maior nos homens por ser capaz de mudar decisões de comportamentos de riscos, principalmente aqueles relacionados a mortes por causas externas (homicídio e acidente de trânsito, por exemplo) (ROSS et al., 2012). Esses resultados confirmam o que a maioria dos trabalhos encontrados na literatura encontraram (MACKENBACH et al., 2008; MONTEZ et al., 2009; MANZELLI, 2014; SILVA, 2014; SANDOVAL; TURRA, 2015; RIBEIRO, 2016; SILVA et al., 2016; KROLL et al., 2017; TURRA et al., 2018). Entretanto, há trabalhos com resultados controversos, em que também não houve diferenciais na mortalidade por escolaridade segundo o sexo ou que esses diferenciais foram maiores nas mulheres (KITAGAWA; HAUSER, 1973; ZAJACOVA, 2006; HUMMER; LARISCY, 2011).

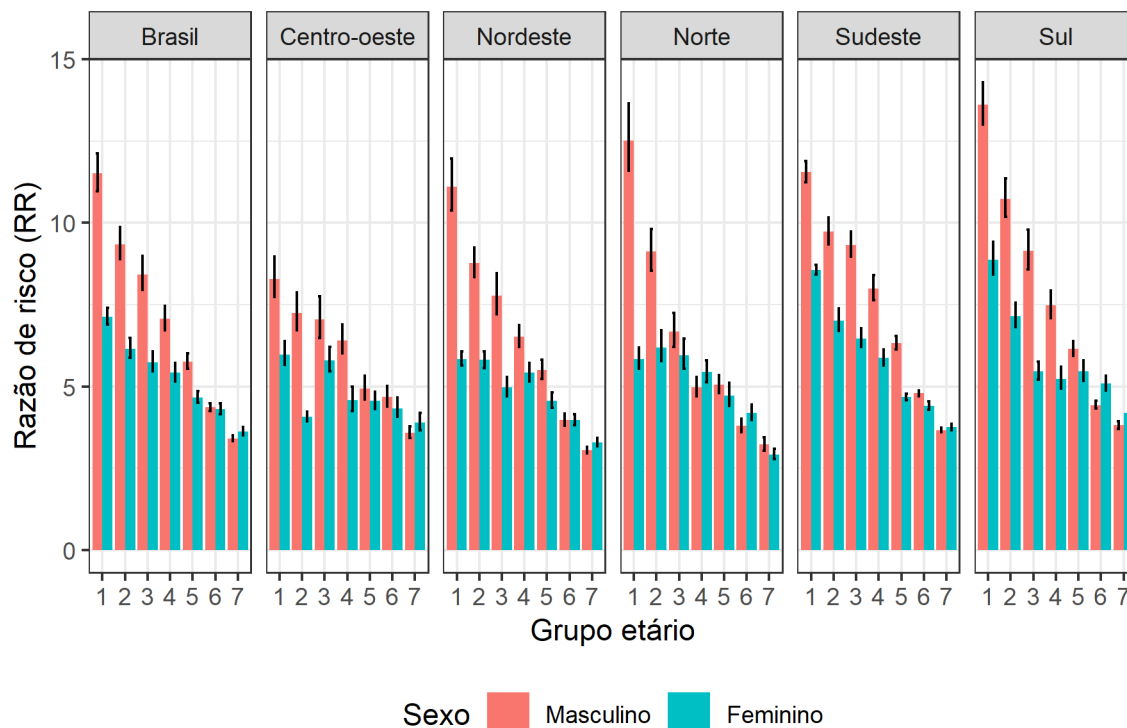
Figura 12: Razão de risco (RR) segundo o sexo por idade e grandes regiões no cenário com correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: Grupo etário = 1 – 25 a 29 anos, 2 – 30 a 34 anos, 3 – 35 a 39 anos, 4 – 40 a 44 anos, 5 – 45 a 49 anos, 6 – 50 a 54 anos, 7 – 55 a 59 anos. RR = (Taxa de mortalidade da baixa escolaridade) / (Taxa de mortalidade da alta escolaridade).

Figura 13: Razão de risco (RR) segundo o sexo por idade e grandes regiões no cenário sem correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

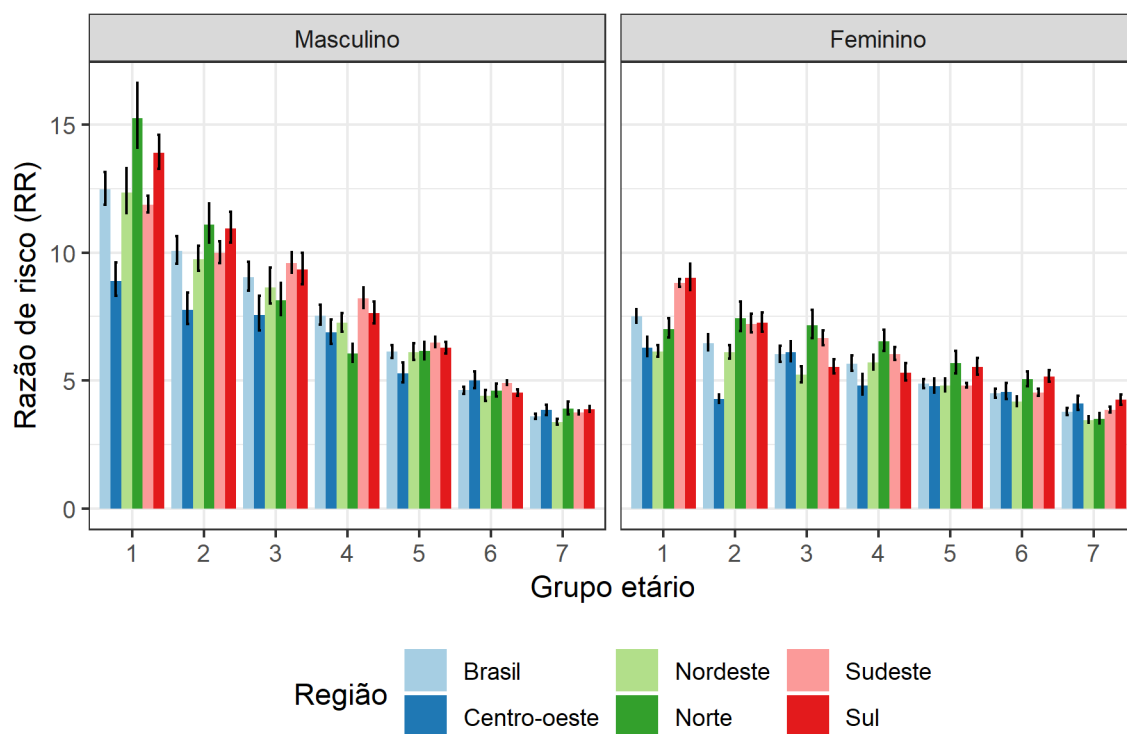
Nota: Grupo etário = 1 – 25 a 29 anos, 2 – 30 a 34 anos, 3 – 35 a 39 anos, 4 – 40 a 44 anos, 5 – 45 a 49 anos, 6 – 50 a 54 anos, 7 – 55 a 59 anos. $RR = (\text{Taxa de mortalidade da baixa escolaridade}) / (\text{Taxa de mortalidade da alta escolaridade})$.

Ainda sobre os diferenciais por sexo, mas analisando as razões entre a probabilidade de morte entre 25 e 59 anos dos óbitos com baixa escolaridade e essa probabilidade nos falecidos com alta escolaridade, constata-se que esse gradiente é maior entre os homens nos dois cenários, o que confirma os resultados obtidos por meio das RRs. Com os óbitos corrigidos, a probabilidade de um homem adulto com baixa escolaridade morrer é 6,17 [5,94 – 6,44] vezes a probabilidade de morte de um homem adulto com alta escolaridade no Brasil. Já para as mulheres adultas esse gradiente é de 5,04 [4,84 – 5,26]. A região de maior diferença entre homens e mulheres em termos de RP é na região Nordeste. Contudo, a vantagem maior que a escolaridade exerce sobre os homens não se verifica na região Norte nos dois cenários estudados (Tabelas A3 e A4 – Apêndice).

Em relação as diferenças regionais na magnitude dos diferenciais por escolaridade segundo a idade, fica evidenciado pelas Figuras 14 (com correção) e 15 (sem correção) que não há um padrão bem definido quanto às regiões que se destacam em termos de magnitude dos diferenciais. O que se observa mais claramente é que as diferenças entre as razões de risco das

regiões tendem a diminuir com a idade. Por exemplo, para os homens na idade de 25 a 29 e 30 a 34 anos (cenário com correção), a RR é maior nas regiões Norte e Sul. Já para as idades de 35 a 39 e 40 a 44 anos, destacam-se as regiões Sudeste e Sul. A partir dos 45 anos quase não se observou diferença ou elas são mínimas entre as regiões. Essa oscilação de padrão do gradiente também ocorre para o sexo masculino e no cenário sem correção de sub-registro.

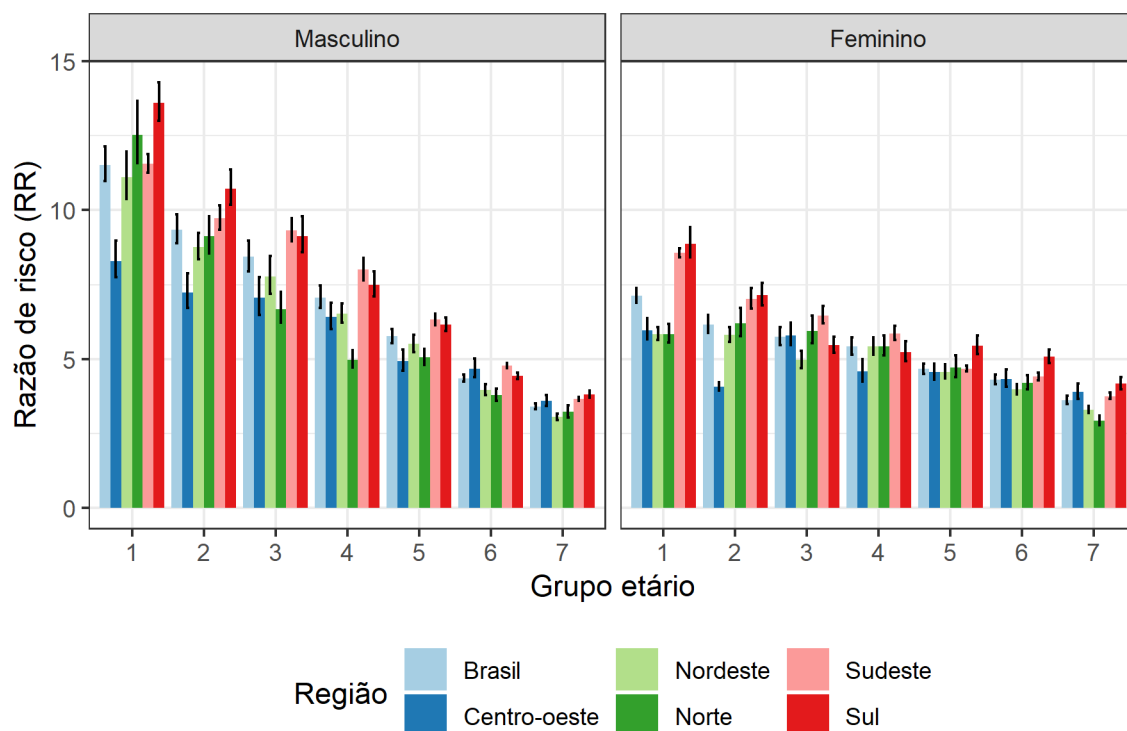
Figura 14: Razão de risco (RR) por regiões, segundo sexo e idade com correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: Grupo etário = 1 – 25 a 29 anos, 2 – 30 a 34 anos, 3 – 35 a 39 anos, 4 – 40 a 44 anos, 5 – 45 a 49 anos, 6 – 50 a 54 anos, 7 – 55 a 59 anos. $RR = (Taxa \text{ de mortalidade da baixa escolaridade}) / (Taxa \text{ de mortalidade da alta escolaridade})$.

Figura 15: Razão de risco (RR) por regiões, segundo sexo e idade sem correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: Grupo etário = 1 – 25 a 29 anos, 2 – 30 a 34 anos, 3 – 35 a 39 anos, 4 – 40 a 44 anos, 5 – 45 a 49 anos, 6 – 50 a 54 anos, 7 – 55 a 59 anos. RR = (Taxa de mortalidade da baixa escolaridade) / (Taxa de mortalidade da alta escolaridade).

Uma maneira de tentar detectar diferenças regionais na magnitude dos diferenciais por escolaridade é pelo indicador resumo 35q25 (ver Figuras A5 e A6 – Apêndice). Por meio da razão de probabilidade (RP) no cenário com correção de sub-registro, verifica-se que praticamente não há diferenças regionais em termos de magnitude dos diferenciais de mortalidade por escolaridade para o sexo masculino. Todas as regiões apresentaram RP médios em torno de 6,00, ou seja, a probabilidade de um indivíduo do sexo masculino com baixa escolaridade vir a falecer é 6,00 vezes a probabilidade dos que possuem alta escolaridade para todas as regiões. No caso de não correção de sub-registros, conclusões equivocadas de que os maiores diferenciais foram nas regiões Sul e Sudeste para o sexo masculino seriam retiradas.

Em relação ao sexo feminino (cenário com correção), existem diferenças, ainda que mínimas, entre as RPs das regiões Sudeste, Sul e Norte em relação as RPs do Nordeste e Centro-oeste. A maior diferença foi entre a região Norte (RP = 5,70; IC[5,36 – 6,61]) e Centro-oeste (RP = 4,68; IC[4,41 – 5,01]).

Esses resultados são totalmente inéditos e contrários aos encontrados na literatura. No trabalho de Manzelli (2014) para Argentina, o autor afirma ter encontrado maiores diferenciais

em áreas mais escolarizadas. No caso do Brasil, Silva (2014) e Silva et al. (2016) argumenta que os maiores diferenciais são observados no Nordeste e Sudeste para os dois sexos. Por outro lado, Turra et al. (2018) encontraram maiores gradientes nas regiões Sudeste e Centro-oeste.

Uma questão que merece ser chamada atenção aqui é a importância de se considerar os intervalos de confiança quando se realiza imputação por algum processo aleatório em trabalhos que visem estimar diferenças de mortalidade por escolaridade, bem como da devida correção de sub-registro dos óbitos. Caso o pesquisador não leve isso em consideração, poderá incorrer em conclusões equivocadas, conforme ficou evidente na análise de cenários.

6 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo estimar as diferenças de mortalidade por escolaridade segundo o sexo, idade e grandes regiões a partir dos dados de óbitos registrados no SIM em 2010. Para isso, algumas etapas metodológicas foram fundamentais.

Primeiro realizou-se uma avaliação da completude da variável escolaridade do SIM. O resultado foi que a chance de o registro de óbito apresentar escolaridade faltante difere segundo variáveis sociodemográficas. Esse resultado foi de grande relevância para o processo de imputação da variável escolaridade. Uma possibilidade de estudo futuro para essa temática seria investigar se essas diferenças na chance de ausência ocorrem ao longo do período de 1996 a 2016, pós mudança da CID-9 para CID-10.

No que se refere à imputação da variável escolaridade dos registros de óbitos, uma hipótese levantada foi que imputar não alteraria a distribuição percentual por idade e sexo e ela foi confirmada. Ressaltou-se, ao longo do capítulo 4, que a proposta aqui não seria esgotar as possibilidades metodológicas para resolução desse problema. Dessa forma, sugere-se que em trabalhos futuros sejam aplicados outros métodos de imputação para que seja possível comparar e avaliar a sensibilidade deles. Além disso, recomenda-se validar as estimativas das variâncias calculadas sob a teoria de IM e, conseqüentemente, os intervalos de confiança construídos para a média da frequência de óbitos em cada idade por região, sexo e escolaridade utilizadas neste estudo. Fay (1996) argumenta que mesmo quando um estimador de variância é consistente, como é o caso da variância proposta por Rubin no contexto de IM, estudos envolvendo simulações são importantes.

No tocante aos principais resultados desse estudo, que são os diferenciais de mortalidade por escolaridade, observou-se que a maioria dos resultados estão coerentes com o encontrado na literatura. Considerando que os resultados no cenário com correção de sub-registro são os definitivos deste trabalho, ficou evidente que o risco de morte nos indivíduos que possuem baixa ou média escolaridade é sempre maior comparado àqueles que possuem alta escolaridade. Ao realizar a correção de sub-registro apenas nos óbitos de menor escolaridade, os gradientes na taxa de mortalidade entre os óbitos com média escolaridade e baixa escolaridade ficaram mais notáveis, principalmente nos homens adultos mais jovens. Quando as probabilidades de morte foram avaliadas, a relação inversa com a escolaridade só não foi observada para as mulheres da região nordeste comparando baixa com média escolaridade. No que diz respeito aos diferenciais por sexo e idade ficou evidenciado que eles são maiores entre os homens e reduzem com idade, sendo essa redução mais brusca para os homens. Um dos resultados que

chamou mais atenção foi que, contrariamente ao que a literatura aponta, não há diferenciais regionais entre a magnitude dos gradientes de mortalidade por escolaridade.

Embora todos os cuidados metodológicos para correção das quantidades de óbitos tenham sido tomados, é importante ter cautela com possíveis inconsistências entre o numerador e o denominador das taxas de mortalidade por escolaridade. Isso porque eles provêm de fontes de dados distintas e, conseqüentemente, a escolaridade é registrada de maneira distinta. No SIM (numerador) a escolaridade é informada por terceiros que forneceram a informação para preenchimento da declaração de óbito. No censo (denominador), a pessoa de referência do domicílio declara a escolaridade dos moradores do domicílio.

O estudo de Turra et al.(2018) traz uma reflexão importante acerca dessa discussão. Os autores utilizaram duas maneiras distintas para obter o denominador das taxas a partir do total populacional da PNAD 2014: uma baseada na estrutura por sexo, idade e escolaridade do censo de 2010; e a outra obtida pela estrutura por idade e sexo do censo 2010 e estrutura de escolaridade das regiões metropolitanas onde as capitais utilizadas no estudo se inserem. O resultado foi que as taxas mudam dependendo da estrutura do denominador. Comparando as taxas calculadas com esses dois denominadores, eles observaram que as maiores diferenças são nos grupos de escolaridade de 12 anos ou mais de estudo e que elas variam com a idade (TURRA et al., 2018). Isso mostra alguma incerteza quanto a verdadeira estrutura populacional por escolaridade no Brasil e, conseqüentemente, dos verdadeiros diferenciais de mortalidade por essa variável.

Outra questão que merece atenção é em relação à correção de sub-registro dos óbitos por escolaridade. Na ausência de estimativas de graus de cobertura por escolaridade, a suposição de que a sub-enumeração dos óbitos ocorre apenas naqueles com baixa escolaridade é razoável. Porém, ficou evidente a necessidade de estudos para estimar a cobertura dos óbitos por essa variável, pois afetam consideravelmente a magnitude dos diferenciais de mortalidade. A quebra de pressupostos (por exemplo, populações fechadas) e as mudanças na forma de mensurar a escolaridade nos censos 2000 e 2010 inviabilizam a aplicação dos métodos de distribuição de morte (como o SEG adj.). Desse modo, para trabalhos futuros, sugere-se considerar a metodologia proposta por Adair e Lopez (2018) que é uma alternativa aos métodos de distribuição de morte livre de suposições ligadas à dinâmica populacional. Essa correção ficou inviável neste trabalho devido à grande carga metodológica e de resultados já apresentados.

Apesar de suas limitações, o presente trabalho fornece uma importante contribuição na busca de uma definição das verdadeiras magnitudes e dos verdadeiros padrões dos diferenciais de mortalidade por escolaridade no Brasil. Talvez a solução para resolver esses problemas de inconsistências entre o numerador e denominador seja a implementação de pesquisas longitudinais prospectivas, em que grupos de indivíduos são acompanhados ao longo do tempo até a ocorrência do desfecho morte. Ou então, realizar o pareamento de pesquisas domiciliares de saúde (como a PNS, por exemplo) com os dados de registro de óbito para análise de dados em painel.

Por fim, espera-se também que esse trabalho venha a alertar aos gestores públicos que medidas para garantir a qualidade do preenchimento das declarações de óbito são extremamente urgentes e relevantes para que se tenha fontes de dados confiáveis para uma gestão da saúde pública com mais eficiência e assertividade.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

ABOUZAHR, C.; BOERMA, T. Health information systems: the foundations of public health. **Bulletin of the World Health Organization**, v. 83, n. 8, p. 578-583, 2005.

ADAIR, T.; LOPEZ, A. D. Estimating the completeness of death registration: An empirical method. **PLOS ONE**, v. 13, n. 5, p. e0197047, 2018. <https://doi.org/10.1371/journal.pone.0197047>.

ALLISON, P. D. **Missing data**. Thousand Oaks, CA: Sage publications, 2001.

ALVES, J. E. D. **A transição demográfica e a janela de oportunidade**. São Paulo: Instituto Fernand Braudel de Economia Mundial, 2008.

ANDRIDGE, R. R.; LITTLE, R. J. A. A Review of Hot Deck Imputation for Survey Non-response. **International statistical review = Revue internationale de statistique**, v. 78, n. 1, p. 40-64, 2010. <https://doi.org/10.1111/j.1751-5823.2010.00103.x>.

BARATA, R. B. **Como e por que as desigualdades sociais fazem mal à saúde**. Rio de Janeiro: Editora Fiocruz, 2009.

BENNETT, N. G.; HORIUCHI, S. Estimating the Completeness of Death Registration in a Closed Population. **Population Index**, v. 47, n. 2, p. 207-221, 1981. <https://doi.org/10.2307/2736447>.

BODNER, T. E. What Improves with Increased Missing Data Imputations? **Structural Equation Modeling: A Multidisciplinary Journal**, v. 15, n. 4, p. 651-675, 2008. <https://doi.org/10.1080/10705510802339072>.

BORGES, G. M. Health transition in Brazil: regional variations and divergence/convergence in mortality. **Cadernos de Saúde Pública**, v. 33, 2017. <http://dx.doi.org/10.1590/0102-311x00080316>.

BOSWORTH, B. Increasing Disparities in Mortality by Socioeconomic Status. **Annual Review of Public Health**, v. 39, n. 1, p. 237-251, 2018. <http://doi.org/10.1146/annurev-publhealth-040617-014615>.

BRITO, F. Transição demográfica e desigualdades sociais no Brasil. **Revista Brasileira de Estudos de População**, v. 25, p. 5-26, 2008. <http://dx.doi.org/10.1590/S0102-30982008000100002>.

CALDWELL, J. C. Cultural and Social Factors Influencing Mortality Levels in Developing Countries. **The ANNALS of the American Academy of Political and Social Science**, v. 510, n. 1, p. 44-59, 1990. <http://doi.org/10.1177/0002716290510001004>.

_____. Health transition: The cultural, social and behavioural determinants of health in the Third World. **Social Science & Medicine**, v. 36, n. 2, p. 125-135, 1993. [https://doi.org/10.1016/0277-9536\(93\)90204-H](https://doi.org/10.1016/0277-9536(93)90204-H).

CAMARGOS, V. P.; CÉSAR, C. C.; CAIAFFA, W. T.; XAVIER, C. C.; PROIETTI, F. A. Imputação múltipla e análise de casos completos em modelos de regressão logística: uma avaliação prática do impacto das perdas em covariáveis. **Cadernos de Saúde Pública**, v. 27, p. 2299-2313, 2011. <http://dx.doi.org/10.1590/S0102-311X2011001200003>.

CARVALHO, J. A. M. D.; RODRÍGUEZ-WONG, L. L. A transição da estrutura etária da população brasileira na primeira metade do século XXI. **Cadernos de Saúde Pública**, v. 24, p. 597-605, 2008. <http://dx.doi.org/10.1590/S0102-311X2008000300013>.

COALE, A. J. The demographic transition: a summary, some lessons, and some observations. In: CHO, L.; KOBAYASHI, K. (Ed.). **Fertility transition of the East Asian populations**. ed. Honolulu: University Press of Hawaii, 1979. cap. 2, p. 9-23.

COALE, A. J. Demographic transition. In: EATWELL, J.; MILGATE, M.; NEWMAN, P. (Ed.). **Social economics**. ed. London: Springer, 1989. cap. 4, p. 16-23.

COHEN, S.; SYME, S. L. **Social support and health**. San Diego, CA, US: Academic Press, 1985.

COLLINS, L. M.; SCHAFER, J. L.; KAM, C.-M. A comparison of inclusive and restrictive strategies in modern missing data procedures. **Psychological Methods**, v. 6, n. 4, p. 330-351, 2001. <https://doi.org/10.1037/1082-989X.6.4.330>.

CORREIA, L. O. D. S.; PADILHA, B. M.; VASCONCELOS, S. M. L. Métodos para avaliar a completude dos dados dos sistemas de informação em saúde do Brasil: uma revisão sistemática. **Ciência & Saúde Coletiva**, v. 19, p. 4467-4478, 2014. <http://dx.doi.org/10.1590/1413-812320141911.02822013>.

COSTA, J. M. B. D. S.; FRIAS, P. G. D. Avaliação da completude das variáveis da declaração de óbitos de menores de um ano residentes em Pernambuco, 1997-2005. **Ciência & Saúde Coletiva**, v. 16, p. 1267-1274, 2011. <http://dx.doi.org/10.1590/S1413-81232011000700059>.

CRIMMINS, E. M. Socioeconomic differentials in mortality and health at the older ages. **Genus**, v. 61, n. 1, p. 163-176, 2005.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the royal statistical society. Series B (methodological)**, p. 1-38, 1977.

ELO, I. T.; PRESTON, S. H. Educational differentials in mortality: United States, 1979–1985. **Social Science & Medicine**, v. 42, n. 1, p. 47-57, 1996. [https://doi.org/10.1016/0277-9536\(95\)00062-3](https://doi.org/10.1016/0277-9536(95)00062-3).

ENDERS, C. K. **Applied missing data analysis**. New York, NY: Guilford press, 2010.

FAY, R. E. Alternative Paradigms for the Analysis of Imputed Survey Data. **Journal of the American Statistical Association**, v. 91, n. 434, p. 490-498, 1996. <https://doi.org/10.1080/01621459.1996.10476909>.

FELIX, J. D.; ZANDONADE, E.; AMORIM, M. H. C.; CASTRO, D. S. D. Avaliação da completude das variáveis epidemiológicas do Sistema de Informação sobre Mortalidade em mulheres com óbitos por câncer de mama na Região Sudeste: Brasil (1998 a 2007). **Ciência & Saúde Coletiva**, v. 17, p. 945-953, 2012. <http://dx.doi.org/10.1590/S1413-81232012000400016>.

FERNANDES, D. M. Estimativas do Grau de Cobertura e da Esperança de Vida para as Unidades da Federação no Brasil entre 2000 e 2010. In: Encontro de Estudos de População da ABEP, 2. 1984, São Paulo, SP. **Anais...** São Paulo, SP: ABEP, 1984.

FILHO, A. M. S.; SOUZA, M. D. F. M. D.; GAZAL-CARVALHO, C.; MALTA, D. C.; ALENCAR, A. P.; SILVA, M. M. A. D.; NETO, O. L. D. M. Análise da mortalidade por homicídios no Brasil. **Epidemiologia e Serviços de Saúde**, v. 16, p. 7-18, 2007. <http://dx.doi.org/10.5123/S1679-49742007000100002>.

FRENK, J.; BOBADILLA, J. L.; STERN, C.; FREJKA, T.; LOZANO, R. Elements for a theory of the health transition. **Health Transition Review**, v. 1, n. 1, p. 21-38, 1991a.

FRENK, J.; FREJKA, T.; BOBADILLA, J. L.; STERN, C.; LOZANO, R.; SEPÚLVEDA, J.; JOSÉ, M. La transición epidemiológica en América Latina. **Boletín de la Oficina Sanitaria Panamericana (OSP)**, v. 111, p. 485-96, 1991b.

GALDINO, A.; SANTANA, V. S.; FERRITE, S. Quality of the record of data on fatal workplace injuries in Brazil. **2017**, v. 51, 2017. <http://dx.doi.org/10.11606/S1518-8787.2017051000064>.

GOMES, M. M. F.; TURRA, C. M. The number of centenarians in Brazil: Indirect estimates based on death certificates. **Demographic Research**, v. 20, n. 20, p. 495-502, 2009. <https://doi.org/10.4054/DemRes.2009.20.20>.

GONZAGA, M. R.; SCHMERTMANN, C. P. Estimating age-and sex-specific mortality rates for small areas with TOPALS regression: an application to Brazil in 2010. **Revista Brasileira de Estudos de População**, v. 33, n. 3, p. 629-652, 2016. <http://dx.doi.org/10.20947/s0102-30982016c0009>.

GRAHAM, J. W.; OLCOWSKI, A. E.; GILREATH, T. D. How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. **Prevention Science**, v. 8, n. 3, p. 206-213, 2007. <https://doi.org/10.1007/s11121-007-0070-9>.

GUEDES, G. R.; SIVIERO, P. C. L.; QUEIROZ, B. L.; MACHADO, C. J. Approximating the educational differences in mortality: demographic indirect techniques. **Cad. Saúde Colet. Rio de Janeiro**, v. 19, n. 2, p. 240-3, 2011.

HARRELL, F. **Regression Modeling Strategies With Applications to Linear Models, Logistic Regression, and Survival Analysis**. New York, NY: Springer, 2001.

HILL, K. Estimating census and death registration completeness. In: Asian and Pacific population forum/East-West Population Institute, East-West Center, 3, 1987: The Asian & Pacific Population Forum. 8-13, 23 p.

HILL, K.; YOU, D.; CHOI, Y. Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data errors. **Demographic Research**, v. 21, n. 9, p. 235-254, 2009. <https://doi.org/10.4054/DemRes.2009.21.9>.

HOUSE, J.; LANDIS, K.; UMBERSON, D. Social relationships and health. **Science**, v. 241, n. 4865, p. 540-545, 1988. <https://doi.org/10.1126/science.3399889>.

HUMMER, R. A.; HERNANDEZ, E. M. The Effect of Educational Attainment on Adult Mortality in the United States*. **Population bulletin**, v. 68, n. 1, p. 1-16, 2013.

HUMMER, R. A.; LARISCY, J. T. Educational attainment and adult mortality. In: (Ed.). **International handbook of adult mortality**. ed.: Springer, 2011. cap., p. 241-261.

KIM, J. K. A Note on Approximate Bayesian Bootstrap Imputation. **Biometrika**, v. 89, n. 2, p. 470-477, 2002.

KINNEY, S. K. Multiple imputation with census data. **Catalogue no. 12-001-XIE**, p. 215, 2012.

KITAGAWA, E. M.; HAUSER, P. M. **A differential mortality in the United States: a study in socioeconomic epidemiology**. Cambridge, MA: Harvard University Press, 1973.

KROLL, L. E.; BAUMERT, J.; BUSCH, M.; SCHEIDT-NAVE, C.; LAMPERT, T.; HOEBEL, J. Estimating social inequality in mortality and life expectancy in absence of register data in Germany Jens Hoebel. **European Journal of Public Health**, v. 27, n. suppl_3, p. cxx189.158-ckx189.158, 2017. <https://doi.org/10.1093/eurpub/ckx189.158>.

KUNST, A. E.; BOS, V.; ANDERSEN, O.; CARDANO, M.; COSTA, G.; HARDING, S.; HEMSTRÖM, Ö.; LAYTE, R.; REGIDOR, E.; REID, A.; SANTANA, P.; VALKONEN, T.; MACKENBACH, J. P. Monitoring of trends in socioeconomic inequalities in mortality: Experiences from a European project. **Demographic Research**, v. S2, p. 229-254, 2004. <https://doi.org/10.4054/DemRes.2004.S2.9>.

KUNST, A. E.; MACKENBACH, J. P. The size of mortality differences associated with educational level in nine industrialized countries. **American Journal of Public Health**, v. 84, n. 6, p. 932-937, 1994.

LAUDERDALE, D. S. Education and survival: Birth cohort, period, and age effects. **Demography**, v. 38, n. 4, p. 551-561, 2001. <https://doi.org/10.1353/dem.2001.0035>.

LERNER, M. Modernization and health: a model of the health transition. In: Annual Meeting of the American Public Health Association, 1973, São Francisco, Califórnia. American Public Health Association, 1973.

LEVY, R. B.; CASTRO, I. R. R. D.; CARDOSO, L. D. O.; TAVARES, L. F.; SARDINHA, L. M. V.; GOMES, F. D. S.; COSTA, A. W. N. D. Consumo e comportamento alimentar entre adolescentes brasileiros: Pesquisa Nacional de Saúde do Escolar (PeNSE), 2009. **Ciência & Saúde Coletiva**, v. 15, p. 3085-3097, 2010.

LIMA, C. R. D. A.; SCHRAMM, J. M. D. A.; COELI, C. M.; SILVA, M. E. M. D. Revisão das dimensões de qualidade dos dados e métodos aplicados na avaliação dos sistemas de informação em saúde. **Cadernos de Saúde Pública**, v. 25, p. 2095-2109, 2009. <http://dx.doi.org/10.1590/S0102-311X2009001000002>.

LITTLE, R. J.; RUBIN, D. B. **Statistical analysis with missing data**. 2nd ed. Hoboken, NJ: John Wiley & Sons, 2002.

LITTLE, R. J. A.; RUBIN, D. **Statistical Analysis with Missing Data**. New York: John Wiley and Sons, 1987.

LOWRY, D.; XIE, Y. **Socioeconomic status and health differentials in China: convergence or divergence at older ages?** Population Studies Center, University of Michigan, 2009.

MACENTE, L. B.; ZANDONADE, E. Avaliação da completude do sistema de informação sobre mortalidade por suicídio na região Sudeste, Brasil, no período de 1996 a 2007. **Jornal Brasileiro de Psiquiatria**, v. 59, p. 173-181, 2010. <http://dx.doi.org/10.1590/S0047-20852010000300002>.

MACKENBACH, J. P. The persistence of health inequalities in modern welfare states: The explanation of a paradox. **Social Science & Medicine**, v. 75, n. 4, p. 761-769, 2012. <https://doi.org/10.1016/j.socscimed.2012.02.031>.

MACKENBACH, J. P.; KULHÁNOVÁ, I.; ARTNIK, B.; BOPP, M.; BORRELL, C.; CLEMENS, T.; COSTA, G.; DIBBEN, C.; KALEDIENE, R.; LUNDBERG, O.; MARTIKAINEN, P.; MENVIELLE, G.; ÖSTERGREN, O.; PROCHORSKAS, R.; RODRÍGUEZ-SANZ, M.; STRAND, B. H.; LOOMAN, C. W. N.; DE GELDER, R. Changes in mortality inequalities over two decades: register based study of European countries. **BMJ**, v. 353, 2016. <https://doi.org/10.1136/bmj.i1732>.

MACKENBACH, J. P.; KULHÁNOVÁ, I.; BOPP, M.; DEBOOSERE, P.; EIKEMO, T. A.; HOFFMANN, R.; KULIK, M. C.; LEINSALU, M.; MARTIKAINEN, P.; MENVIELLE, G.; REGIDOR, E.; WOJTYNIAK, B.; ÖSTERGREN, O.; LUNDBERG, O. Variations in the relation between education and cause-specific mortality in 19 European populations: A test of the “fundamental causes” theory of social inequalities in health. **Social Science & Medicine**, v. 127, p. 51-62, 2015. <https://doi.org/10.1016/j.socscimed.2014.05.021>.

MACKENBACH, J. P.; KUNST, A. E.; CAVELAARS, A. E. J. M.; GROENHOF, F.; GEURTS, J. J. M. Socioeconomic inequalities in morbidity and mortality in western Europe. **The Lancet**, v. 349, n. 9066, p. 1655-1659, 1997. [https://doi.org/10.1016/S0140-6736\(96\)07226-1](https://doi.org/10.1016/S0140-6736(96)07226-1).

MACKENBACH, J. P.; STIRBU, I.; ROSKAM, A.-J. R.; SCHAAP, M. M.; MENVIELLE, G.; LEINSALU, M.; KUNST, A. E. Socioeconomic Inequalities in Health in 22 European Countries. **New England Journal of Medicine**, v. 358, n. 23, p. 2468-2481, 2008. <https://doi.org/10.1056/NEJMsa0707519>.

MACKENBACH, J. P.; VALVERDE, J. R.; ARTNIK, B.; BOPP, M.; BRØNNUM-HANSEN, H.; DEBOOSERE, P.; KALEDIENE, R.; KOVÁCS, K.; LEINSALU, M.; MARTIKAINEN, P.; MENVIELLE, G.; REGIDOR, E.; RYCHTAŘÍKOVÁ, J.; RODRIGUEZ-SANZ, M.; VINEIS, P.; WHITE, C.; WOJTYNIAK, B.; HU, Y.; NUSSELDER, W. J. Trends in health inequalities in 27 European countries. **Proceedings of the National Academy of Sciences**, 2018. <https://doi.org/10.1073/pnas.1800028115>.

MALTA, D. C.; FRANÇA, E.; ABREU, D. X. D.; OLIVEIRA, H. D.; MONTEIRO, R. A.; SARDINHA, L. M. V.; DUARTE, E. C.; SILVA, G. A. E. Atualização da lista de causas de mortes evitáveis (5 a 74 anos de idade) por intervenções do Sistema Único de Saúde do Brasil. **Epidemiologia e Serviços de Saúde**, v. 20, p. 409-412, 2011. <https://doi.org/10.5123/S1679-49742011000300016>.

MANZELLI, H. Educational attainment and adult mortality differentials in Argentina. **Revista Latinoamericana de Población**, v. 8, n. 14, p. 129-163, 2014.

MONTEZ, J. K.; HAYWARD, M. D.; BROWN, D. C.; HUMMER, R. A. Why Is the Educational Gradient of Mortality Steeper for Men? **The Journals of Gerontology Series B: Psychological Sciences and Social Sciences**, v. 64B, n. 5, p. 625-634, 2009. <https://doi.org/10.1093/geronb/gbp013>.

MONTEZ, J. K.; HUMMER, R. A.; HAYWARD, M. D. Educational Attainment and Adult Mortality in the United States: A Systematic Analysis of Functional Form. **Demography**, v. 49, n. 1, p. 315-336, 2012. <https://doi.org/10.1007/s13524-011-0082-8>.

MONTEZ, J. K.; HUMMER, R. A.; HAYWARD, M. D.; WOO, H.; ROGERS, R. G. Trends in the Educational Gradient of U.S. Adult Mortality from 1986 to 2006 by Race, Gender, and Age Group. **Research on aging**, v. 33, n. 2, p. 145-171, 2011. <https://doi.org/10.1177/0164027510392388>.

MS - MINISTÉRIO DA SAÚDE. Coordenação Geral de Informações e Análise Epidemiológica – CGIAE. **Sistema de Informações sobre Mortalidade – SIM: Consolidação da base de dados de 2011**. 2011. 12 p.

MULLER, A. Education, income inequality, and mortality: a multiple regression analysis. **BMJ**, v. 324, n. 7328, p. 23, 2002. <https://doi.org/10.1136/bmj.324.7328.23>.

NATIONAL CENTER FOR HEALTH STATISTICS. **Health, United States, 2011: With Special Feature on Socioeconomic Status and Health**: (US). National Center for Health Statistics (US), Hyattsville (MD), 2012.

NOTESTEIN, F. W. Population: the long view. In: SCHULTZ, T. (Ed.). **Food for the world**. ed. Chicago, Illinois: University of Chicago Press, 1945. cap., p. 36-57.

NUNES, A.; SANTOS, J. R. S.; BARATA, R. B.; VIANNA, S. M. Medindo as desigualdades em saúde no Brasil. **Brasília: Organização Pan-Americana da Saúde/Instituto de Pesquisa Econômica Aplicada**, 2001.

NUNES, L. N. **Métodos de imputação de dados aplicados na área da saúde. 2007. 120 p.** 2007. p. - Tese (Doutorado em Epidemiologia)–Universidade Federal do Rio Grande do Sul, Porto Alegre,

NUNES, L. N.; KLÜCK, M. M.; FACHEL, J. M. G. Uso da imputação múltipla de dados faltantes: uma simulação utilizando dados epidemiológicos. **Cadernos de Saúde Pública**, v. 25, p. 268-278, 2009. <http://dx.doi.org/10.1590/S0102-311X2009000200005>.

_____. Comparação de métodos de imputação única e múltipla usando como exemplo um modelo de risco para mortalidade cirúrgica. **Revista Brasileira de Epidemiologia**, v. 13, p. 596-606, 2010. <http://dx.doi.org/10.1590/S1415-790X2010000400005>.

OLIVEIRA, M. M. D.; LATORRE, M. D. R. D. D. O.; TANAKA, L. F.; CURADO, M. P. Simulação e comparação de técnicas de correção de dados incompletos de idade para o cálculo de taxas de incidência. **Cad. Saúde Pública**, v. 34, n. 6, p. e00140717, 2018. <http://dx.doi.org/10.1590/0102-311X00140717>.

OMRAN, A. R. The Epidemiologic Transition: A Theory of the Epidemiology of Population Change. **The Milbank Quarterly**, v. 49, n. 4, p. 509-538, 1971. <https://doi.org/10.1111/j.1468-0009.2005.00398.x>.

_____. The epidemiologic transition theory revisited thirty years later. **World health statistics quarterly**, v. 51, n. 2-4, p. 99-119, 1998.

ORTEGA, A. **Tablas de mortalidad**. Naciones Unidas Comisión Económica para América Latina y el Caribe (CEPAL). 1987

PAES, N. A. Avaliação da cobertura dos registros de óbitos dos estados brasileiros em 2000. **Revista de Saúde Pública**, v. 39, p. 882-890, 2005. <http://dx.doi.org/10.1590/S0034-89102005000600003>

PAES, N. A.; ALBUQUERQUE, M. E. E. Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras. **Revista de Saúde Pública**, v. 33, p. 33-43, 1999. <http://dx.doi.org/10.1590/S0034-89101999000100006>.

PAIVA, P. D. T. A.; WAJNMAN, S. Das causas às consequências econômicas da transição demográfica no Brasil. **Revista brasileira de estudos populacionais**, v. 22, n. 2, p. 13-15, 2005.

PÉREZ, E. R. **Estimativas de mortalidade adulta feminina por nível de escolaridade no Brasil**. 2010. 166 p. Tese (Doutorado em Demografia) - CEDEPLAR, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

PÉREZ, E. R.; TURRA, C. M. Desigualdade social na mortalidade no Brasil: diferenciais por escolaridade entre mulheres adultas. In: Encontro Nacional de Estudos Populacionais 16., 2008, Caxambu, MG. **Anais...** Caxambu, MG: ABEP, 2008.

PO, J. Y. T.; SUBRAMANIAN, S. V. Mortality Burden and Socioeconomic Status in India. **PLoS ONE**, v. 6, n. 2, p. e16844, 2011. <http://dx.doi.org/10.1371/journal.pone.0016844>.

PRATA, P. R. A transição epidemiológica no Brasil. **Cadernos de Saúde Pública**, v. 8, p. 168-175, 1992. <http://dx.doi.org/10.1590/S0102-311X1992000200008>.

PRESTON, S.; HEUVELINE, P.; GUILLOT, M. **Demography: Measuring and Modeling Population Processes**. Wiley-Blackwell, 2001.

PRESTON, S. H.; ELO, I. T.; STEWART, Q. Effects of Age Misreporting on Mortality Estimates at Older Ages. **Population Studies**, v. 53, n. 2, p. 165-177, 1999.

PRESTON, S. H.; TAUBMAN, P. Socioeconomic Differences in Adult Mortality and Health Status. In: MARTIN, L. G.; PRESTON, S. H. (Ed.). **Demography of Aging**. ed. Washington, D.C.: National Academy Press, 1994. cap., p. 279-318.

QUEIROZ, B. L. Estimativas do Grau de Cobertura e da Esperança de Vida para as Unidades da Federação no Brasil entre 2000 e 2010. In: Encontro de Estudos de População da ABEP, 18. 2012, Águas de Lindóia, SP. **Anais...** Águas de Lindóia, SP: ABEP, 2012.

QUEIROZ, B. L.; FREIRE, F. H. M. D. A.; GONZAGA, M. R.; LIMA, E. E. C. D. Estimativas do grau de cobertura e da mortalidade adulta (45q15) para as unidades da federação no Brasil entre 1980 e 2010. **Revista Brasileira de Epidemiologia**, v. 20, p. 21-33, 2017. <http://dx.doi.org/10.1590/1980-5497201700050003>.

QUEIROZ, B. L.; SAWYER, D. O. O que os dados de mortalidade do Censo de 2010 podem nos dizer. **Rev Bras Estud Popul**, v. 29, n. 2, p. 225-238, 2012.

RIBEIRO, M. M. **Mortalidade adulta por níveis de escolaridade no estado e município de São Paulo: uma proposta de estimação a partir do Censo Demográfico de 2010**. 2016. 123 p. Tese (Doutorado em Demografia) - CEDEPLAR, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

ROMERO, D. E.; CUNHA, C. B. D. Avaliação da qualidade das variáveis sócio-econômicas e demográficas dos óbitos de crianças menores de um ano registrados no Sistema de Informações sobre Mortalidade do Brasil (1996/2001). **Cadernos de Saúde Pública**, v. 22, p. 673-681, 2006. <http://dx.doi.org/10.1590/S0102-311X2006000300022>.

ROSS, C. E.; MASTERS, R. K.; HUMMER, R. A. Education and the Gender Gaps in Health and Mortality. **Demography**, v. 49, n. 4, p. 1157-1183, 2012. <https://doi.org/10.1007/s13524-012-0130-z>.

ROSTRON, B. L.; BOIES, J. L.; ARIAS, E. Education reporting and classification on death certificates in the United States. **Vital and health statistics. Series 2, Data evaluation and methods research**, n. 151, p. 1-21, 2010.

RUBIN, D. B. Inference and missing data. **Biometrika**, v. 63, n. 3, p. 581-592, 1976. <https://doi.org/10.1093/biomet/63.3.581>.

RUBIN, D. B. Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. In: Proceedings of the survey research methods section of the American Statistical Association, 1978: American Statistical Association. 20-34 p.

RUBIN, D. B. **Multiple Imputation for Nonresponse in Surveys**. 1987.

_____. Multiple Imputation after 18+ Years. **Journal of the American Statistical Association**, v. 91, n. 434, p. 473-489, 1996. <https://doi.org/10.1080/01621459.1996.10476908>.

RUBIN, D. B.; SCHENKER, N. Multiple Imputation for Interval Estimation From Simple Random Samples With Ignorable Nonresponse. **Journal of the American Statistical Association**, v. 81, n. 394, p. 366-374, 1986. <https://doi.org/10.2307/2289225>.

SAIKIA, N.; BHAT, P. M. Factors Affecting Adult Mortality in India: An Analysis of National Family Health Surveys of 1992-93 and 1998-99 (NFHS I and II). **Demography India**, v. 37, n. 2, p. 291-302, 2008.

SANDOVAL, M. H.; TURRA, C. M. El gradiente educativo en la mortalidad adulta en Chile. **2015**, n. 17, p. 29, 2015.

SCHAFER, J. L. **Analysis of incomplete multivariate data**. Chapman and Hall/CRC, 1997.

SCHAFER, J. L.; GRAHAM, J. W. Missing data: Our view of the state of the art. **Psychological Methods**, v. 7, n. 2, p. 147-177, 2002. <https://doi.org/10.1037/1082-989X.7.2.147>.

SCHRAMM, J. M. D. A.; OLIVEIRA, A. F. D.; LEITE, I. D. C.; VALENTE, J. G.; GADELHA, Â. M. J.; PORTELA, M. C.; CAMPOS, M. R. Transição epidemiológica e o estudo de carga de doença no Brasil. **Ciência & Saúde Coletiva**, v. 9, p. 897-908, 2004. <http://dx.doi.org/10.1590/S1413-81232004000400011>.

SILVA JÚNIOR, W. P.; FREIRE, F. H. M. D. A. Determinantes sociodemográficos da incompletude dos dados de escolaridade do óbito adulto no Brasil em 2010. In: Encontro de Estudos de População da ABEP, 2018, Poços de Caldas, MG. **Anais...** Poços de Caldas, MG: ABEP, 2018.

SILVA, L. E. D. **Diferenciais de mortalidade adulta por nível de escolaridade no Brasil e Regiões**. 2014. 111 p. Dissertação (Mestrado em Demografia) - DDCA, Universidade Federal do Rio Grande do Norte, Natal, 2014.

SILVA, L. E. D.; FREIRE, F. H. M. D. A.; PEREIRA, R. H. M. Diferenciais de mortalidade por escolaridade da população adulta brasileira, em 2010. **Cadernos de Saúde Pública**, v. 32, 2016. <http://dx.doi.org/10.1590/0102-311X00019815>.

SORLIE, P. D.; JOHNSON, N. J. Validity of Education Information on the Death Certificate. **Epidemiology**, v. 7, n. 4, p. 437-439, 1996.

STUQUE, C. D. O.; CORDEIRO, J. A.; CURY, P. M. Avaliação dos erros ou falhas de preenchimento dos atestados de óbito feitos pelos clínicos e pelos patologistas. **Jornal**

Brasileiro de Patologia e Medicina Laboratorial, v. 39, p. 361-364, 2003. <http://dx.doi.org/10.1590/S1676-24442003000400014>.

SUBRAMANIAN, S. V.; NANDY, S.; IRVING, M.; GORDON, D.; LAMBERT, H.; DAVEY SMITH, G. The Mortality Divide in India: The Differential Contributions of Gender, Caste, and Standard of Living Across the Life Course. **American Journal of Public Health**, v. 96, n. 5, p. 818-825, 2006. <http://dx.doi.org/10.2105/AJPH.2004.060103>.

TRAVASSOS, C.; OLIVEIRA, E. X. G. D.; VIACAVA, F. Desigualdades geográficas e sociais no acesso aos serviços de saúde no Brasil: 1998 e 2003. **Ciência & Saúde Coletiva**, v. 11, p. 975-986, 2006. <http://dx.doi.org/10.1590/S1413-81232006000400019>.

TURRA, C. M.; RENTERIA, E.; GUIMARÃES, R. The Effect of Changes in Educational Composition on Adult Female Mortality in Brazil. **Research on aging**, v. 38, n. 3, p. 283-298, 2016. <https://doi.org/10.1177/0164027515620245>.

TURRA, C. M.; RIBEIRO, M. M.; PINTO, C. C. D. X. Padrões de mortalidade por escolaridade no Brasil: evidências a partir do Sistema de Informação sobre Mortalidade. In: Encontro Nacional de Estudos Populacionais 2018, Poços de Caldas, MG. **Anais...** Poços de Caldas, MG: ABEP, 2018.

VALKONEN, T. Adult mortality and level of education: a comparison of six countries. In: FOX, J. (Ed.). **Health Inequalities in European Countries**. ed. Aldershot: Gower Publishing Company, 1989. cap., p. 142-162.

VASCONCELOS, A. M. N.; GOMES, M. M. F. Transição demográfica: a experiência brasileira. **Epidemiologia e Serviços de Saúde**, v. 21, p. 539-548, 2012. <http://dx.doi.org/10.5123/S1679-49742012000400003>

WILLIAMS, D. R. Socioeconomic Differentials in Health: A Review and Redirection. **Social Psychology Quarterly**, v. 53, n. 2, p. 81-99, 1990. <https://doi.org/10.2307/2786672>.

WOOD, C. H.; CARVALHO, J. A. M. D. A demografia da desigualdade no Brasil. In: (Ed.). **A demografia da desigualdade no Brasil**. ed., 1994. cap., p.

ZAJACOVA, A. Education, gender, and mortality: Does schooling have the same effect on mortality for men and women in the US? **Social Science & Medicine**, v. 63, n. 8, p. 2176-2190, 2006. <https://doi.org/10.1016/j.socscimed.2006.04.031>.

ZAJACOVA, A.; HUMMER, R. A. Gender differences in education effects on all-cause mortality for white and black adults in the United States. **Social Science & Medicine**, v. 69, n. 4, p. 529-537, 2009. <https://doi.org/10.1016/j.socscimed.2009.06.028>.

APÊNDICE

Tabela A1: Lista de variáveis disponíveis nos microdados e no dicionário dos dados de óbito do SIM em 2010 segundo a utilização no estudo e a condição de preenchimento.

Variável	Descrição	Utilização / Condição de preenchimento
D_R	Não consta no dicionário	Não utilizada / Todos os registros são <i>missing</i>
numerodo	Número da DO	Não utilizada / Preenchida para todos
codinst	Não consta no dicionário	Não utilizada / Preenchida para todos
origem	Não consta no dicionário	Não utilizada / Preenchida para todos
tipobito	Tipo de Óbito	Não utilizada / Preenchida para todos
dtobito	Data do Óbito	Não utilizada / Preenchida para todos
horaobito	Hora do falecimento	Não utilizada / Preenchida para todos
natural	Naturalidade	Não utilizada / Preenchida para todos
dtnasc	Data de Nascimento	Não utilizada / Preenchida para todos
idade	Idade	Utilizada / Preenchida para todos
sexo	Sexo	Utilizada / Preenchida para todos
racacor	Raça/cor	Utilizada / Preenchida para todos
estciv	Estado Civil	Utilizada / Preenchida para todos
esc	Escolaridade	Utilizada / Preenchida para todos
ocup	Ocupação, conforme CBO	Não utilizada / Preenchida para todos
codmunres	Código do Município de Residência do Óbito	Não utilizada / Preenchida para todos
codbaires	Código do bairro de residência	Não utilizada / Preenchida para todos
lococor	Local de ocorrência do óbito	Utilizada / Preenchida para todos
codestab	Código do estabelecimento	Não utilizada / Preenchida para todos
codmunocor	Código do Município de Ocorrência do Óbito	Não utilizada / Preenchida para todos
codbaiocor	Código do bairro de ocorrência	Não utilizada / Preenchida para todos
idademae	Idade da mãe	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
escmae	Escolaridade da mãe	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
ocupmae	Ocupação da mãe	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
qtdfilvivo	Número de filhos vivos	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
qtdfilmort	Número de filhos mortos	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
gravidez	Tipo de gravidez	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
gestacao	Semana de gestação	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
parto	Tipo de parto	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
obitoparto	Morte em relação ao parto	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
peso	Peso ao nascer	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
numerodn	Número da DN	Não utilizada / Preenchida apenas para óbitos fetais ou para óbitos menores de 1 ano
obitograv	Morte durante a gravidez	Não utilizada / Preenchida apenas para mulheres
obitopuerp	Morte durante o puerpério	Não utilizada / Preenchida apenas para mulheres
assistmed	Indica de houve Assistência médica	Não utilizada / Todos os registros são <i>missing</i>
exame	Indica de houve Exame complementar	Não utilizada / Preenchida para todos
cirurgia	Indica de houve Cirurgia	Não utilizada / Preenchida para todos
necropsia	Indica de houve necropsia	Não utilizada / Preenchida para todos
linhaa	Linha A do atestado CID10	Não utilizada / Preenchida para todos
linhab	Linha B do atestado CID10	Não utilizada / Preenchida para todos
linhac	Linha C do atestado CID10	Não utilizada / Preenchida para todos
linhad	Linha D do atestado CID10	Não utilizada / Preenchida para todos
linhaii	Linha II do atestado CID10	Não utilizada / Preenchida para todos
causabas	Causa Básica CID10	Utilizada / Preenchida para todos

Variável	Descrição	Utilização / Condição de preenchimento
dtatestado	Data do Atestado	Não utilizada / Preenchida para todos
circobito	Indica o tipo de acidente	Não utilizada / Preenchida apenas para mortes violentas
acidtrab	Indica se foi acidente de trabalho	Não utilizada / Preenchida apenas para mortes violentas
fonte	Fonte de informação	Não utilizada / Preenchida apenas para mortes violentas
tppos	Óbito investigado	Não utilizada / Preenchida apenas para mortes violentas
dtinvestig	Data de investigação	Não utilizada / Preenchida apenas para mortes violentas
causabas_o	Causa básica original, a primeira informação que entra no sistema.	Não utilizada / Preenchida para todos
dtcadastro	Data de cadastro do registro no sistema	Não utilizada / Preenchida para todos
atestante	Indica se o médico que assina atendeu o paciente	Não utilizada / Preenchida para todos
fonteinv	Fonte de investigação	Não utilizada / Preenchida apenas para mortes violentas
dtrecebim	Data de recebimento no nível central, data da última atualização do registro	Não utilizada / Preenchida para todos
ufinform	Código da UF que informou o registro	Não utilizada / Todos os registros são <i>missing</i>
cb_pre	Causa selecionada sem re-seleção (novo SCB)	Não utilizada / Preenchida para todos
morteparto	Não consta no dicionário	Não utilizada / Todos os registros são <i>missing</i>
dtcadinf	Não consta no dicionário	Não utilizada / Todos os registros são <i>missing</i>
tpobitocor	Não consta no dicionário	Não utilizada / Preenchida para todos
dtcadinv	Não consta no dicionário	Não utilizada / Preenchida para todos
tpassina	Não consta nos microdados	Não utilizada / Todos os registros são <i>missing</i>

Fonte: Ministério da Saúde / Datasus

Tabela A2: Descrição, categorização original proposta no dicionário, a codificação nos microdados, a codificação publicada em consulta no site do Datasus e a categorização final utilizada nas variáveis do SIM em 2010 utilizadas no capítulo 3.

Nome nos microdados	Nome utilizado	Descrição	Codificação dicionário	Codificação no microdados	Codificação do site*	Codificação final
NA**	Região	Região de residência do falecido	NA	NA	1: Norte	1: Norte
					2: Nordeste	2: Nordeste
					3: Sudeste	3: Sudeste
					4: Sul	4: Sul
					5: Centro Oeste	5: Centro Oeste
sexo	Sexo	Sexo do falecido	1: Masculino	1: Masculino	1: Masculino	1: Masculino
			2: Feminino	2: Feminino	2: Feminino	2: Feminino
			0: Ignorado	9: Ignorado	9: Ignorado	9: Ignorado
idade	Idade	Idade do falecido	Idade, composto de dois subcampos. O primeiro, de 1 dígito, indica a unidade da idade. O segundo, de dois dígitos, indica a quantidade de unidades: 1: Horas, o segundo subcampo varia de 01 a 23 2: Dias, o segundo subcampo varia de 01 a 29	Igual ao dicionário	6: 25 a 29 anos	6: 25 a 29 anos
					7: 30 a 34 anos	7: 30 a 34 anos
					8: 35 a 39 anos	8: 35 a 39 anos

Nome nos microdados	Nome utilizado	Descrição	Codificação dicionário	Codificação no microdados	Codificação do site*	Codificação final
			3: Meses, o segundo subcampo varia de 01 a 11 4: Anos, o segundo subcampo varia de 00 a 99 5: Anos (mais de 100 anos), o segundo subcampo varia de 0 a 99 5: Anos (mais de 100 anos), o segundo subcampo varia de 0 a 99 0: Idade ignorada, o segundo subcampo é 00;		9: 40 a 44 anos 10: 45 a 49 anos 11: 50 a 54 anos 12: 55 a 59 anos	9: 40 a 44 anos 10: 45 a 49 anos 11: 50 a 54 anos 12: 55 a 59 anos
				000, 999 e em branco: Ignorado	999: Ignorado (000, 999, em branco)	999: Ignorado (000, 999, em branco)
racacor	Cor/raça	Cor ou raça do falecido	1:Branca	1:Branca	1:Branca	1: Branca
			2:Preta	2:Preta	2:Preta	2: Preta
			3:Amarela	3:Amarela	3:Amarela	4: Parda
			4: Parda	4: Parda	4: Parda	6: Amarela ou indígena
			5: Indígena	5: Indígena Em branco	5: Indígena Em branco: Ignorado	9: Ignorado (Em branco)
estciv	Estado civil	Estado civil do falecido	1: Solteiro	1: Solteiro	1: Solteiro	1: Solteiro
			2: Casado	2: Casado	2: Casado	2: Casado
			3: Viúvo	3: Viúvo	3: Viúvo	3: Viúvo
			4: Separado judicialmente	4: Separado judicialmente	4: Separado judicialmente	6: Divorciado/Outros (4 e 5)
			5: "não há nome para a categoria"	5: "não há nome para a categoria"	5: Outro	
			9: Ignorado	9: Ignorado	9: Ignorado	9: Ignorado
lococor	Local de ocorrência	Local de ocorrência do óbito	1: Hospital			1: Hospital
			2: Outro estab saúde			2: Outro estabelecimento de saúde
			3: Domicílio	Igual ao dicionário	Igual ao dicionário	3: Domicílio
			4: Via Pública			4: Via pública
			5: Outros			5: Outros
			9: Ignorado			9: Ignorado
causabas	Causa básica de morte (em capítulos)	Causa Básica de morte conforme CID10	Causa básica, conforme a Classificação Internacional de Doença (CID), 10a. Revisão	Igual ao dicionário	Igual ao dicionário	2: II. Neoplasias (tumores) 9: IX. Doenças do aparelho circulatório 10: X. Doenças do aparelho respiratório 20: XX. Causas externas 99: Outros
esc	Escolaridade de	Escolaridade e do falecido (em anos de estudo concluídos)	1: Nenhuma	1: Nenhum	1: Nenhum	1: Nenhum
			2: 1 a 3 anos	2: 1 a 3 anos	2: 1 a 3 anos	2: 1 a 3 anos
			3: 4 a 7 anos	3: 4 a 7 anos	3: 4 a 7 anos	3: 4 a 7 anos
			4: 8 a 11 anos	4: 8 a 11 anos	4: 8 a 11 anos	4: 8 a 11 anos
			5: 12 e mais	5: 12 anos e mais	5: 12 anos e mais	5: 12 anos e mais

Nome nos microdados	Nome utilizado	Descrição	Codificação dicionário	Codificação no microdados	Codificação do site*	Codificação final
			9: Ignorado	9: Ignorado	9, 0 e em branco: Ignorado	9: Ignorado (9, 0 e em branco)
				0: "não há nome para a categoria"		
				Em branco		
NA	Escolaridade de (3 categorias)	Escolaridade e com recategorização em baixa, média e alta escolaridade e a partir da variável "esc"	NA	NA	NA	1, 2, 3 => 1: baixa 4 => 2: média 5 => 3: alta 9 => 9: Ignorado
NA	Escolaridade de faltante	Indica se a escolaridade e foi missing ou não a partir da variável "esc"	NA	NA	NA	1,2,3,4,5 => 0 : Não 9 => 1: Sim

Fonte: Ministério da Saúde / Datasus

* Site do Datasus com as categorizações: <http://tabnet.datasus.gov.br/cgi/defthtm.exe?sim/cnv/obt10uf.def>

**NA = não se aplica

Tabela A3: Percentual de pessoas de 25 a 59 anos com a variável “nível de instrução” não preenchida segundo grandes regiões, sexo e idade nos dados da amostra do Censo 2010.

Idade	Brasil			Norte			Nordeste		
	Feminino	Masculino	total	Feminino	Masculino	total	Feminino	Masculino	total
25 a 29	0,51%	0,44%	0,48%	0,54%	0,38%	0,46%	0,46%	0,33%	0,40%
30 a 34	0,43%	0,34%	0,39%	0,45%	0,29%	0,37%	0,36%	0,24%	0,30%
35 a 39	0,35%	0,29%	0,32%	0,32%	0,23%	0,28%	0,27%	0,20%	0,24%
40 a 44	0,28%	0,25%	0,26%	0,26%	0,18%	0,22%	0,21%	0,14%	0,17%
45 a 49	0,24%	0,21%	0,22%	0,23%	0,21%	0,22%	0,16%	0,13%	0,15%
50 a 54	0,19%	0,17%	0,18%	0,18%	0,13%	0,16%	0,13%	0,11%	0,12%
55 a 59	0,12%	0,14%	0,13%	0,12%	0,14%	0,13%	0,08%	0,07%	0,08%
Total Geral	0,33%	0,29%	0,31%	0,35%	0,25%	0,30%	0,27%	0,20%	0,24%
Idade	Sudeste			Sul			Centro-Oeste		
	Feminino	Masculino	total	Feminino	Masculino	total	Feminino	Masculino	total
25 a 29	0,59%	0,56%	0,58%	0,39%	0,34%	0,36%	0,50%	0,42%	0,46%
30 a 34	0,51%	0,45%	0,48%	0,29%	0,25%	0,27%	0,45%	0,31%	0,38%
35 a 39	0,45%	0,38%	0,42%	0,25%	0,22%	0,23%	0,26%	0,29%	0,27%
40 a 44	0,36%	0,36%	0,36%	0,20%	0,15%	0,17%	0,30%	0,21%	0,26%
45 a 49	0,30%	0,29%	0,30%	0,18%	0,11%	0,15%	0,22%	0,18%	0,20%
50 a 54	0,25%	0,23%	0,24%	0,10%	0,13%	0,12%	0,22%	0,12%	0,17%
55 a 59	0,16%	0,20%	0,18%	0,10%	0,08%	0,09%	0,09%	0,16%	0,12%
Total Geral	0,40%	0,38%	0,39%	0,23%	0,20%	0,21%	0,32%	0,27%	0,30%

Fonte: Microdados do Censo Demográfico/IBGE - 2010

Tabela A4: Taxas específicas de mortalidade (TEM) por 100.000 habitantes para o sexo masculino por idade, segundo escolaridade e grandes regiões, 2010.

Região	Idade	Escolaridade								
		Baixa			Média			Alta		
		TEM (média)	TEM (LI)	TEM (LS)	TEM (média)	TEM (LI)	TEM (LS)	TEM (média)	TEM (LI)	TEM (LS)
Brasil	25 a 29	576,78	568,70	584,86	379,91	367,03	392,79	46,26	43,29	49,24
	30 a 34	539,76	533,31	546,21	384,13	370,45	397,80	53,58	50,12	57,03
	35 a 39	560,79	554,19	567,39	413,68	398,70	428,65	62,05	57,55	66,56
	40 a 44	680,05	672,40	687,70	526,68	506,48	546,88	90,12	84,42	95,82
	45 a 49	889,64	881,54	897,74	776,03	754,45	797,62	145,00	137,79	152,22
	50 a 54	1195,98	1185,75	1206,21	1180,65	1150,48	1210,81	258,66	249,07	268,25
	55 a 59	1556,47	1545,06	1567,87	1725,61	1681,05	1770,16	431,10	416,47	445,73
Centro-oeste	25 a 29	534,36	516,38	552,33	434,08	407,80	460,36	60,11	53,69	66,54
	30 a 34	540,85	529,83	551,87	388,79	365,92	411,66	69,66	62,74	76,59
	35 a 39	551,09	539,65	562,52	391,70	363,15	420,25	72,84	64,84	80,85
	40 a 44	667,05	655,88	678,22	505,53	479,47	531,59	96,99	88,76	105,22
	45 a 49	869,90	853,61	886,19	743,93	697,18	790,68	164,62	149,63	179,61
	50 a 54	1158,38	1143,60	1173,15	1096,18	1045,49	1146,86	231,07	212,91	249,23
	55 a 59	1542,02	1523,59	1560,45	1614,97	1539,51	1690,44	400,28	375,12	425,44
Nordeste	25 a 29	602,12	596,62	607,62	356,81	344,29	369,33	48,77	44,88	52,67
	30 a 34	557,09	552,60	561,57	377,77	364,89	390,65	57,16	53,79	60,53
	35 a 39	558,53	553,87	563,20	402,12	386,10	418,14	64,64	58,90	70,38
	40 a 44	641,95	635,83	648,07	512,97	487,84	538,10	88,54	83,27	93,80
	45 a 49	817,81	812,13	823,49	727,86	702,68	753,03	133,65	125,58	141,72
	50 a 54	1054,06	1044,79	1063,34	1098,26	1055,82	1140,70	238,94	225,72	252,16
	55 a 59	1351,72	1342,53	1360,92	1554,29	1499,45	1609,14	398,73	382,62	414,83
Norte	25 a 29	637,32	627,83	646,81	406,49	392,42	420,57	41,82	37,77	45,87
	30 a 34	587,90	581,04	594,76	433,39	419,25	447,53	52,98	48,71	57,25
	35 a 39	562,44	555,06	569,82	403,05	384,01	422,09	69,15	62,93	75,36
	40 a 44	649,09	640,07	658,10	495,92	471,09	520,75	107,18	99,48	114,88
	45 a 49	836,47	824,75	848,18	683,28	641,83	724,73	135,95	126,70	145,20
	50 a 54	1078,49	1065,63	1091,35	1026,70	982,62	1070,79	233,83	218,15	249,52
	55 a 59	1331,93	1318,30	1345,57	1556,89	1497,21	1616,57	339,68	314,80	364,57
Sudeste	25 a 29	520,47	512,72	528,22	398,44	387,57	409,31	43,85	41,99	45,71
	30 a 34	507,89	501,24	514,55	397,23	386,24	408,22	50,81	48,01	53,61
	35 a 39	566,99	560,54	573,44	447,37	436,10	458,65	59,16	56,03	62,29
	40 a 44	729,10	721,43	736,77	560,91	543,42	578,40	88,74	83,57	93,91
	45 a 49	969,44	963,06	975,81	858,11	845,49	870,72	149,23	143,56	154,91
	50 a 54	1339,51	1329,55	1349,46	1308,45	1284,97	1331,93	272,37	265,42	279,32
	55 a 59	1740,49	1730,87	1750,10	1865,10	1834,08	1896,12	462,56	451,17	473,96
Sul	25 a 29	625,85	615,99	635,70	323,87	312,31	335,44	45,06	42,23	47,90
	30 a 34	540,89	532,47	549,32	327,61	310,04	345,18	49,39	45,92	52,85
	35 a 39	553,84	545,39	562,29	347,57	331,73	363,42	59,37	54,55	64,20
	40 a 44	660,26	651,98	668,54	468,21	450,21	486,21	86,37	80,49	92,26
	45 a 49	865,94	854,57	877,32	645,78	619,70	671,87	137,94	131,01	144,86
	50 a 54	1150,15	1140,76	1159,54	976,46	950,23	1002,70	254,17	245,41	262,93
	55 a 59	1542,18	1526,50	1557,86	1552,31	1492,22	1612,41	396,35	379,62	413,07

Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: Cenário com correção de sub-registro

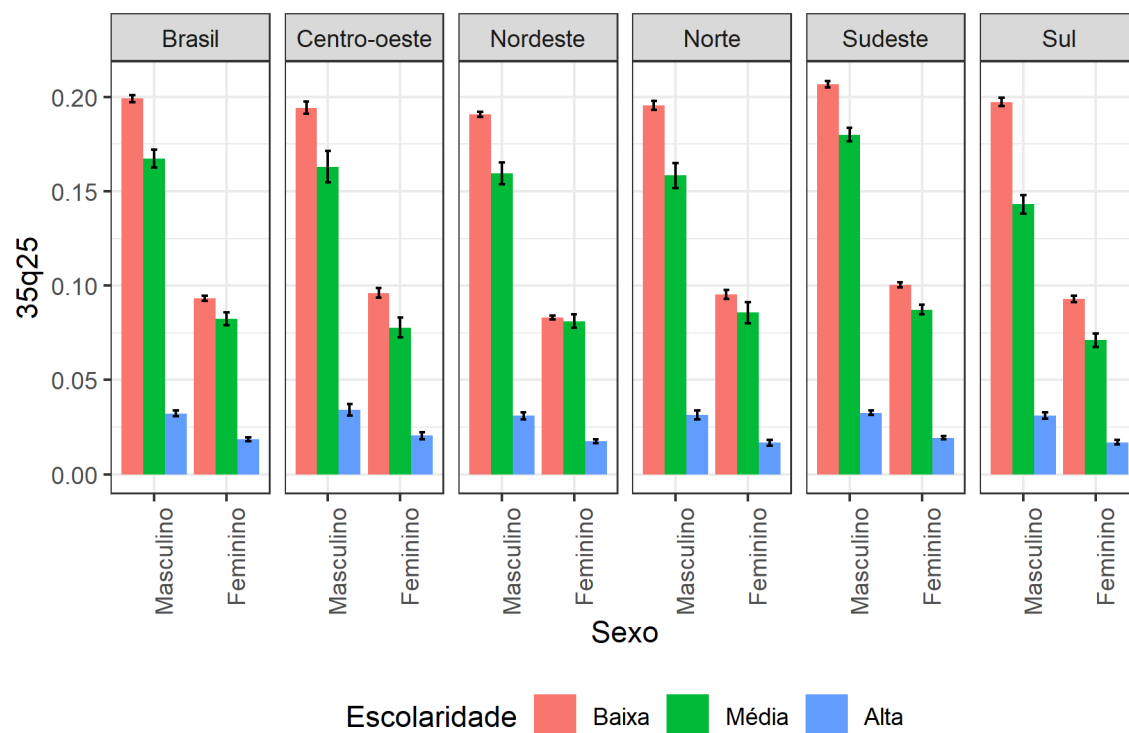
Tabela A5: Taxas específicas de mortalidade (TEM) por 100.000 habitantes para o sexo feminino por idade, segundo escolaridade e grandes regiões, 2010.

Região	Idade	Escolaridade								
		Baixa			Média			Alta		
		TEM (média)	TEM (LI)	TEM (LS)	TEM (média)	TEM (LI)	TEM (LS)	TEM (média)	TEM (LI)	TEM (LS)
Brasil	25 a 29	158,20	153,28	163,11	126,80	120,55	133,06	21,07	19,67	22,48
	30 a 34	183,72	179,03	188,41	159,21	151,24	167,17	28,40	26,29	30,50
	35 a 39	222,86	218,34	227,38	194,67	185,32	204,03	36,99	34,30	39,69
	40 a 44	313,43	308,49	318,37	263,19	251,74	274,65	55,31	51,55	59,08
	45 a 49	453,11	446,66	459,56	392,40	375,37	409,44	92,91	88,16	97,67
	50 a 54	626,04	619,59	632,49	583,00	562,98	603,03	139,01	132,29	145,72
	55 a 59	856,01	848,51	863,50	782,34	752,85	811,83	226,56	216,30	236,81
Centro-oeste	25 a 29	163,60	155,22	171,99	117,73	107,23	128,22	26,02	23,13	28,91
	30 a 34	190,07	183,39	196,75	164,51	152,89	176,13	44,37	41,22	47,53
	35 a 39	226,13	217,32	234,94	202,47	187,23	217,71	37,04	33,24	40,85
	40 a 44	329,18	319,87	338,49	269,56	253,60	285,53	68,32	60,88	75,76
	45 a 49	448,08	437,91	458,24	341,59	315,94	367,23	93,46	85,98	100,94
	50 a 54	662,88	649,56	676,21	524,46	489,33	559,59	145,26	132,61	157,92
	55 a 59	904,57	891,98	917,15	809,43	753,44	865,43	220,09	202,53	237,64
Nordeste	25 a 29	139,46	135,92	143,00	111,49	105,81	117,17	22,70	21,27	24,13
	30 a 34	160,88	157,15	164,61	144,95	137,69	152,20	26,33	24,57	28,08
	35 a 39	191,81	188,50	195,11	184,59	176,74	192,45	36,68	33,88	39,47
	40 a 44	282,24	278,59	285,89	249,80	237,46	262,14	49,50	46,28	52,72
	45 a 49	411,61	406,59	416,63	372,94	354,82	391,06	85,58	80,07	91,09
	50 a 54	550,98	546,08	555,88	628,63	602,96	654,30	131,67	124,84	138,49
	55 a 59	760,75	755,37	766,13	811,06	772,84	849,28	219,25	209,50	229,00
Norte	25 a 29	154,91	149,04	160,78	139,80	131,70	147,89	22,06	20,04	24,09
	30 a 34	194,89	187,50	202,28	182,92	170,37	195,46	26,16	23,21	29,12
	35 a 39	249,55	243,68	255,43	179,56	167,30	191,81	34,86	31,35	38,37
	40 a 44	309,18	299,95	318,40	294,64	271,51	317,77	47,29	43,00	51,59
	45 a 49	456,34	447,96	464,73	384,08	363,73	404,43	80,32	72,76	87,87
	50 a 54	638,98	625,83	652,13	610,71	563,48	657,94	126,56	116,85	136,28
	55 a 59	811,40	798,64	824,17	854,19	807,09	901,29	230,59	214,24	246,94
Sudeste	25 a 29	169,33	164,55	174,10	136,97	132,10	141,83	19,22	18,35	20,08
	30 a 34	197,66	193,99	201,33	164,33	158,47	170,18	27,37	25,50	29,24
	35 a 39	244,45	240,21	248,68	203,61	195,32	211,89	36,72	34,46	38,98
	40 a 44	340,72	336,64	344,80	279,23	270,88	287,58	56,44	53,42	59,46
	45 a 49	490,23	483,49	496,96	429,51	413,96	445,06	101,71	98,26	105,16
	50 a 54	672,23	666,97	677,49	613,42	599,15	627,70	148,20	142,57	153,84
	55 a 59	916,82	909,30	924,34	805,49	783,39	827,59	237,05	228,47	245,63
Sul	25 a 29	189,93	183,15	196,70	123,62	115,60	131,65	21,07	19,13	23,01
	30 a 34	195,87	188,70	203,04	151,87	140,72	163,01	26,98	24,59	29,36
	35 a 39	218,58	213,47	223,69	185,68	175,12	196,24	39,42	36,50	42,33
	40 a 44	300,68	294,91	306,45	219,79	206,73	232,85	56,56	51,84	61,28
	45 a 49	435,66	429,73	441,60	331,06	315,39	346,73	78,66	73,06	84,26
	50 a 54	612,84	605,94	619,73	462,74	444,50	480,97	118,61	112,10	125,13
	55 a 59	855,98	848,83	863,12	654,06	624,15	683,96	201,41	189,85	212,98

Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

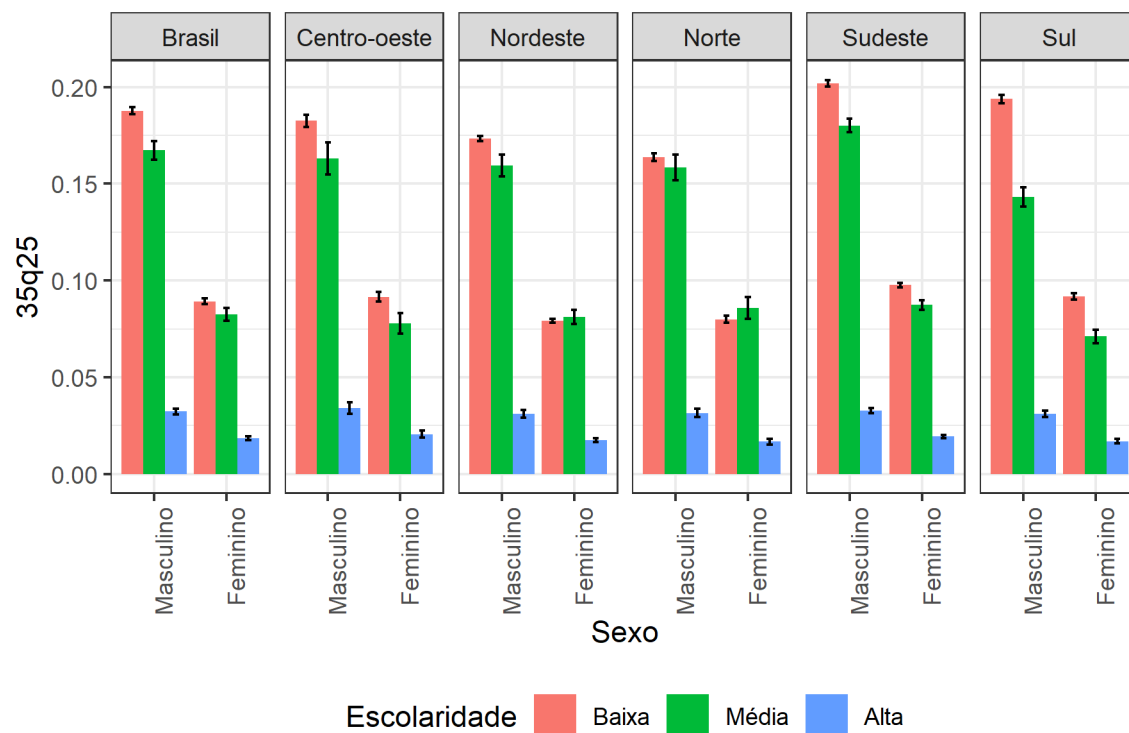
Nota: Cenário com correção de sub-registro

Figura A1: Probabilidade de morte dos 25 aos 59 anos (35q25) por sexo, segundo as regiões com correção de sub-registro, 2010.



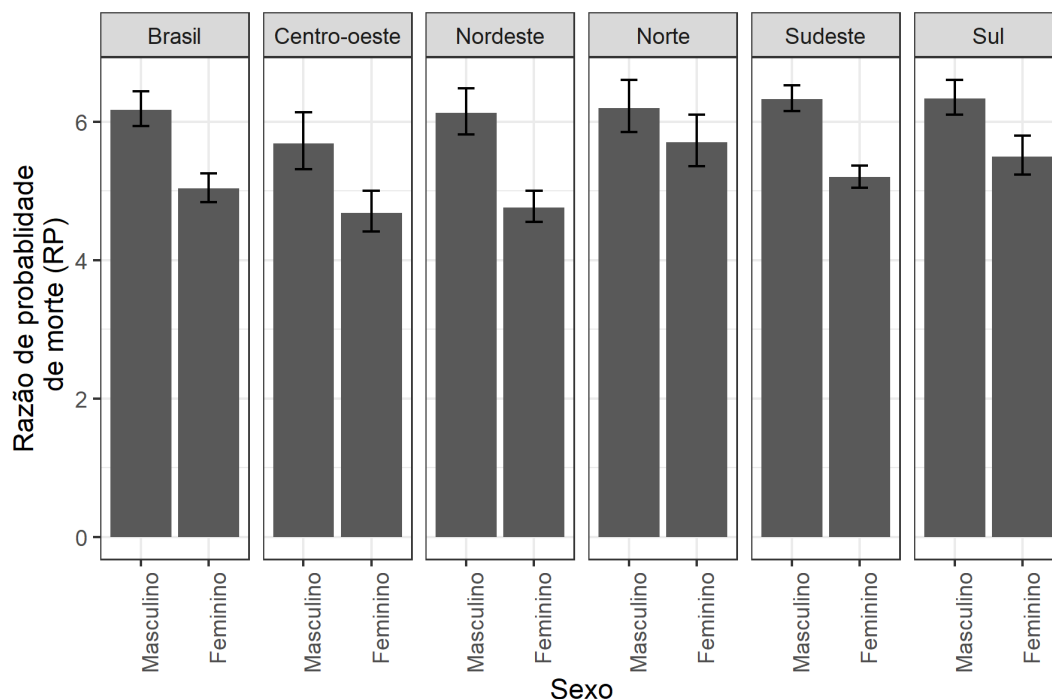
Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Figura A2: Probabilidade de morte dos 25 aos 59 anos (35q25) por sexo, segundo as regiões sem correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

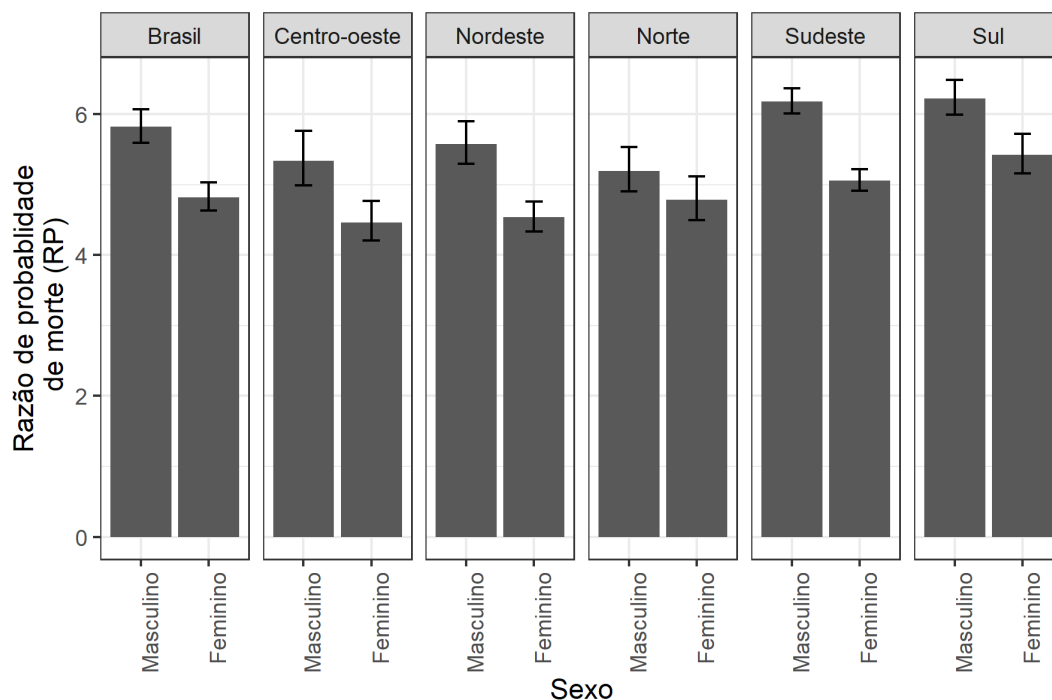
Figura A3: Razão de probabilidade (RP) por sexo, segundo as regiões com correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: RP = (Probabilidade de morte de 25 a 60 da baixa escolaridade) / (Probabilidade de morte de 25 a 60 da alta escolaridade).

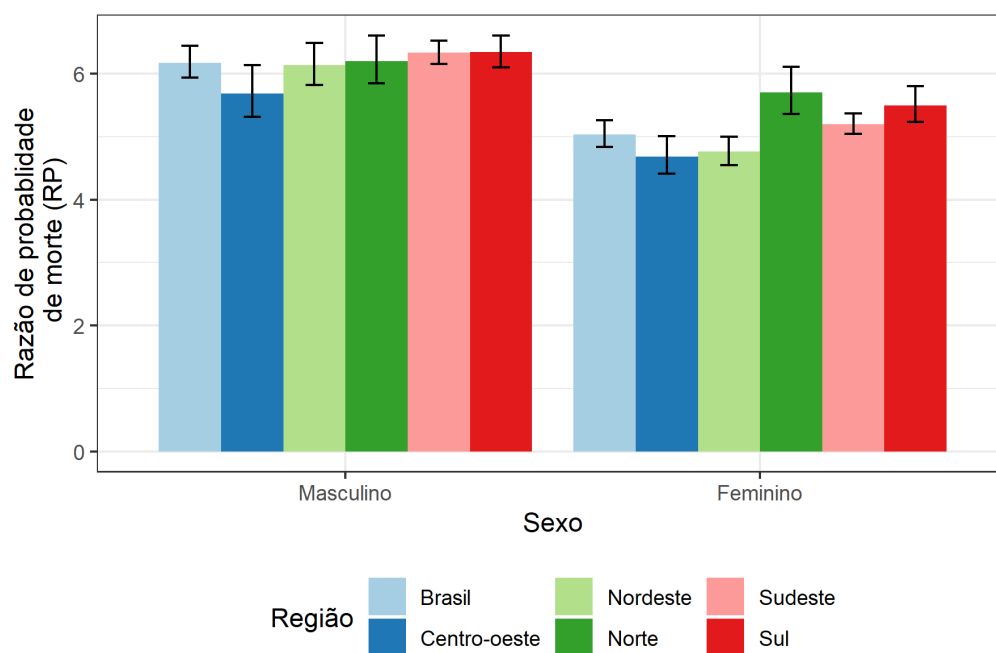
Figura A4: Razão de probabilidade (RP) por sexo, segundo as regiões sem correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: RP = (Probabilidade de morte de 25 a 60 da baixa escolaridade) / (Probabilidade de morte de 25 a 60 da alta escolaridade).

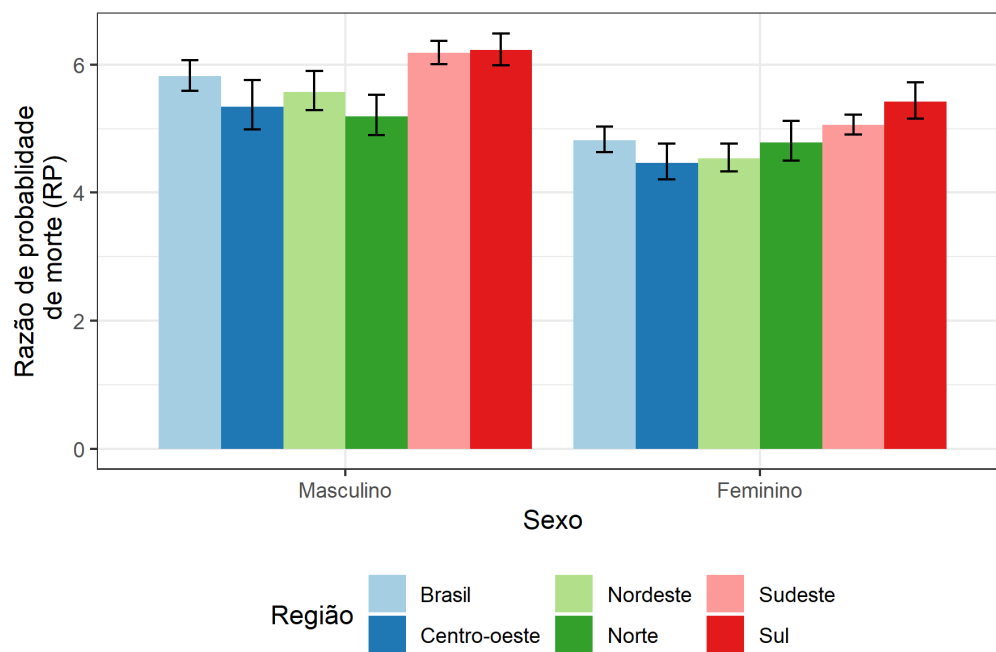
Figura A5: Razão de probabilidade (RP) por região, segundo o sexo com correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: $RP = (\text{Probabilidade de morte de 25 a 60 da baixa escolaridade}) / (\text{Probabilidade de morte de 25 a 60 da alta escolaridade})$.

Figura A6: Razão de probabilidade (RP) por região, segundo o sexo sem correção de sub-registro, 2010.



Fonte: SIM - Sistema de Informação sobre Mortalidade, 2010; IBGE - Censo demográfico, 2010.

Nota: $RP = (\text{Probabilidade de morte de 25 a 60 da baixa escolaridade}) / (\text{Probabilidade de morte de 25 a 60 da alta escolaridade})$.